ENHANCING MODELLING OF LAND-USE AND LAND-COVER CHANGE AND ITS IMPACT ON SURFACE WATER QUALITY IN THE BANGWEULU SUB-CATCHMENT, ZAMBIA

By

MISHECK LESA CHUNDU

A thesis submitted to the University of Zambia in fulfillment of the requirements for the Degree of Doctor of Philosophy in Integrated Water Resources Management.

THE UNIVERSITY OF ZAMBIA

School of Mines, Department of Geology,

Integrated Water Resources Management Centre

Lusaka.

2025

DECLARATION

I, Misheck Lesa Chundu, do hereby declare that this thesis is my original work and that it has not been previously submitted for a degree, diploma, or other qualification at this or any other University. Any incorporated published work or material from other works has been properly referenced and acknowledged.

Misheck Lesa Chundu

Signature of Author: ……. Date: ……

ABSTRACT

Wetlands are among the most productive natural ecosystems in the world, providing essential ecosystem services such as water that benefits human health, supporting aquatic ecosystems, and facilitating economic activities. However, a 64% to 71% of wetlands have been lost globally since 1900, primarily due to changes in land use and land cover (LULC). The Bangweulu Wetland System (BWS) in Zambia faces similar challenges, combined with a lack of comprehensive literature regarding LULC changes and their impacts on surface water quality. Traditional methods of LULC classification can be complex and diverse, but non-parametric approaches, such as Machine Learning (ML), have demonstrated greater accuracy. Different ML models possess distinct strengths and weaknesses, and combining multiple models has the potential to enhance the accuracy of LULC classification. Monitoring various water quality parameters in open water bodies presents significant challenges, resulting in gaps in available data. While the Water Quality Index (WQI) integrates various water quality parameters, its spatial application has not yet been thoroughly evaluated. Although, advancements in Remote Sensing and GIS technology provide wider spatial data coverage, there is a lack of comprehensive literature regarding the impact of LULC on WQI. Additionally, most existing assessment methods have limitations, as they overlook the spatial distribution and proximity of LULC pollution sources to the water body areas of interest. Therefore, the specific objectives of this study were; (i) to ensemble a superiorly hybrid machine learning model for enhanced accuracy of modelling LULC changes, (ii) to rapidly evaluate the water quality variability, and (iii) to investigate the influence of LULC on water quality. To classify LULC, six ML models were employed: Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), Artificial Neural Network (ANN), Random Forest (RF), and K-Nearest Neighbour (KNN). The study analysed Landsat 8 images from 2020 and Landsat 5 images from 1990, 2000, and 2010 using QGIS. To establish an integrated remote sensing approach for monitoring of water quality, the study utilised Sentinel 2 data alongside on-site and laboratory water quality measurements. This integration facilitated the transformation of water quality parameter maps into WQI maps, which were then combined to present an overall assessment of water quality in the BWS lakes. Additionally, a parametric Weighted Inverse Distance Function (WIDF) was applied to determine the contamination effective contribution area (Aec) for each LULC. This analysis utilised the classified Landsat 2020 image, field water quality data and a 30m Digital Elevation Model (DEM). A multiple regression analysis was employed to explore the relationship between Aec and specific water quality parameters, as well as the WQI. Results revealed that four models—SVM, NB, DT, and KNN—outperformed the other models. Consequently, a hybrid model, referred to as the Quad model, was developed by integrating the outputs of these four models.

This Quad model showcased superior performance compared to individual models, achieving Kappa Index scores of 0.87, 0.72, 0.84, and 0.87 for the years 1990, 2000, 2010, and 2020, respectively. The analysis of LULC changes from 1990 to 2020 indicated a yearly decline of -1.17% in forest coverage, -1.01% in grassland, and -0.12% in water bodies. In contrast, built-up areas and cropland increased at rates of 1.70% and 2.70%, respectively. Water quality assessments showed that the mean WQI from on-site and laboratory data was 34.948, while remotely sensed data yielded a mean WQI of 40.633. Both on-site/lab and remote sensing methods indicated that the concentration of water quality parameters in the Bangweulu Wetland lakes is lower (better) than the local and international recommended limits, with a calculated WQI falling within the 'Good' category. This suggests that the water is generally fresh, clean, and suitable for various uses, including ecological preservation, agriculture, aquaculture, recreation, industrial applications, and human consumption. The study also highlighted significant correlations between LULC and water quality parameters. Turbidity, TDS, iron (Fe^{2+}), and EC exhibited strong correlations with specific LULC types, particularly built-up and forested areas. Conversely, parameters such as potassium, sodium, chloride, and calcium showed weak correlations with LULC. The WQI itself demonstrated a reasonable correlation with LULC (R^2 = 0.649). The findings underscore the consistent growth of cropland and built-up areas from 1990 to 2020, alongside a reduction in forest cover and grassland. Although the water body experienced a gradual decrease over this period, the decline was minimal. Long-term monitoring will be essential for evaluating the success of interventions, guiding conservation efforts, and determining whether the reduction in water bodies is a sustained trend or a short-term phenomenon. This information is crucial for developing sustainable LULC policies, identifying hotspots of potential water quality degradation, and targeting areas for restoration efforts, with significant implications for future management practices.

DEDICATION

To my beloved wife, Lillian Mambepa Chundu, whose unwavering support and love have been the foundation of my success. Your strength and encouragement have inspired me every step of the way.

To my wonderful daughters, Moriah and Shekinah Lesa Chundu, may you always strive for excellence and embrace hard work as a path to achieving your dreams. Remember, the dedication and perseverance I have shown in this journey are the same qualities I hope you will embody in your endeavors.

To my mother, Elizabeth Chanda, for her steadfast love and support. I deeply miss my late father, Raymond Lesa, who would have been proud of my achievements. His memories inspire me to work hard and strive for excellence in all I do.

ACKNOWLEDGEMENTS

I am deeply grateful to the OR Tambo Africa Research Chair Initiative for awarding me a remarkable scholarship to pursue my PhD studies at the University of Zambia, without which this research work would not have been feasible.

I am also indebted to my two supervisors, Prof. Kawawa Banda and Prof. Imasiku Nyambe, for their constant availability and invaluable guidance, suggestions, and advice they provided throughout this research project.

Furthermore, I am grateful to Ms. Ingrid Mugamya for her significant contributions in handling administrative and logistical matters efficiently.

Most importantly, I express my heartfelt gratitude to the Almighty God for bestowing upon me the mental fortitude and resilience to persevere throughout this research endeavour.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

Aec	Effective Contribution Area
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARC	Annual Rates Change
BWS	Bangweulu Wetland System
Ca^{2+}	Calcium
Cl^{-}	Chloride
DEM	Digital Elevation Model
DT	Decision Tree
EC	Electrical Conductivity
EO	Earth Observation
ETM+	Enhanced Thematic Mapper Plus
Fe^{2+}	Iron
GIS	Geographical information systems
GMA	Game Management Area
K^{+}	Potassium
KI	Kappa Index
KNN	K-Nearest Neighbour
LULC	Land-Use/Land cover
ML	Machine Learning
MNDWI	Modified Normalized Difference Water Index
Na^{+}	Sodium
NDSI	Normalized Difference Salinity Index
NB	Naive Bayes
NIR	Near-Infrared
NDTI	Normalized Difference Turbid Index
NDWI	Normalized Difference Water Index
OA	Overall Accuracy
OLIS/TIRS	Operational Land Imager/Thermal Infrared Sensor
pH	Potential for Hydrogen
R^2	R-Squared
RF	Random Forest
RS	Remotely Sensed
SVM	Support Vector Machine

TDS	Total Dissolved Solid
TM	Thematic mapper
USGS	United States Geological Survey
WARMA	Water Resources Management Authority
WIDF	Weighted inverse distance function
WQI	Water Quality Index
ZABS	Zambia Bureau of Standards
ZEMA	Zambia Environmental Management Agency

CHAPTER ONE: INTRODUCTION

This chapter provides an introduction of the study, including the background information, problem statement, study objectives, research questions, significance of the research, conceptual framework, and an outline of the thesis.

1.1 Background

Global wetlands are distinct ecosystems where the water table is typically at or near the surface, or the land is flooded, either permanently for years or decades (such as tidal flats and areas near rivers, swamps, or lakes) or seasonally for weeks or months, with static or flowing, fresh, brackish, or shallow marine water (less than six meters deep at low tide) (Semeniuk and Semeniuk, 1995; Stuip *et al*., 2002; Kumar and Kanaujia, 2018). Ramsar Convention on Wetlands (2018) established that global wetlands cover approximately 12.1 million square kilometers, with 54% permanently inundated and 46% seasonally flooded. The extended presence of water creates anoxic (oxygen-free) conditions, particularly in the soils, which promotes the growth of specialised hydrophytic (water-loving) plants and the development of characteristic hydric soils (Morales-Olmedo et al., 2015).

Wetlands provide a wide range of economic, social, environmental, and cultural benefits that are collectively referred to as ecosystem services (de Groot *et al.,* 2018; Delle Grazie and Gill, 2022). Zedler and Kercher (2005) highlights the immense value and importance of wetland ecosystems, which despite covering only 1.5% of the Earth's surface, provide approximately 40% of global ecosystem services. These services span a wide range, including provisioning (food, freshwater, fiber, fuel, genetic resources, medicinal and ornamental resources), regulating (air quality, climate, extreme events, water flows, waste treatment, erosion prevention, soil fertility, pollination), habitat/support (gene pool protection, lifecycle maintenance), and cultural (recreation, tourism, spiritual, aesthetic) benefits. The disproportionately high value of wetlands underscores their crucial role in supporting human well-being and the overall health of the planet, emphasizing the need for their conservation and sustainable management (Zedler and Kercher, 2005; De Araujo Barbosa *et al.*, 2015). It has been established that wetlands are among the world's most biologically diverse and productive natural ecosystems that provide vital ecosystem services such as clean water to people (Damm, 2022).

Wetlands are distinguished from terrestrial landforms by their unique vegetation adapted to the distinctive hydric soil conditions, and there differences are greatly due to regional and local differences in soils, topography, climate, water, vegetation, and other factors, such as human disturbance (Thorp and Covich, 2015). Wetlands are mainly classified into two types: inland

wetlands and coastal wetlands (López-Calderón and Riosmena-Rodríguez, 2016). Inland wetlands are generally herbaceous plant-dominated marshes and wet grasslands, shrub-dominated swamps, and tree-dominated wooded swamps such as lakes, oxbow lakes, riverine, waterlogged, and river/streams. Whereas, coastal wetlands are wetlands found in coastal watersheds like sand/beach, intertidal mudflats, saltmarsh, coral reefs, creeks, and lagoons (Kumar and Kanaujia, 2018).

Africa's wetlands are a diverse and distinct, reflecting the continent's varied environmental conditions. From the world's longest rivers to its largest freshwater bodies, these wetland ecosystems are vital for biodiversity and the well-being of local communities. Despite their importance, African wetlands have received relatively less scientific attention compared to other ecosystems, such as forests or wetlands in other regions (Stephenson *et al*., 2020; Simaika et al., 2021). This underscores the need for greater research and conservation efforts to understand and protect these unique and invaluable natural resources. The limited scientific attention on African wetlands hinders several African countries' ability to report on the condition of these ecosystems and the services they provide to local communities, as mandated by the Ramsar Convention. Furthermore, monitoring changes over time to assess whether Ramsar-designated wetland sites have improved, remained stable, or deteriorated has become a daunting task due to the scarcity of data and research on these ecosystems. This lack of information hampers effective conservation and management efforts, ultimately compromising the long-term health and sustainability of these vital ecosystems (Gardner *et al.*, 2015; Davidson *et al.*, 2019; Stephenson *et al*., 2020). The Ramsar Convention, established in 1971, is the global agreement for the protection of wetlands, particularly those of international importance. It was the first modern instrument aimed at halting global wetlands loss and conserving them through wise use and management (Farrier and Tucker, 2000; Gardner, Okuno and Pritchard, 2023). There are over 2,000 Ramsar sites globally, protecting more than 250 million hectares of wetlands, with over 250 of these sites located in Africa (Kingsford *et al.*, 2021).

Zambia, having ratified the Ramsar Convention in 1991, has designated eight wetlands as internationally important, including the Kafue Flats, Bangweulu Swamps, Barotse Floodplains, Luangwa Floodplains, Busanga Swamps, Lukanga Swamps, Lake Mweru-wa-Ntipa, and Lake Tanganyika (Ministry of Lands, Natural Resources and Environmental and Protection, 2015). These Ramsar wetland sites make up approximately 20% (about 15 million hectares) of Zambia's total land area (WWF Zambia, 2018; Leitão *et al.*, 2019; Stephenson *et al.*, 2020). Gaining a better knowledge of the wetland's LULC changes and their influence on water quality has become essential issues in geography, ecology, and sustainable science (Zhang *et al.*, 2017; Msofe *et at.*,

2019; Zhang *et al.*, 2019). Therefore, in developing countries like Zambia, where the majority of people rely on natural resources for their livelihoods (Ministry of Lands Natural Resources and Environmental Protection, 2015), understanding LULC change dynamics and their effects on water quality, as well as the general status of the available open water bodies, is critical for sustainable wetlands resource management.

Zambia's wetlands contribute to a wide range of economic development by supporting numerous economic sectors such as tourism, agriculture, fishing, and forestry (WWF Zambia, 2018). However, they face significant threats due to wetlands changes caused by both natural and anthropogenic or human-induced factors (Ministry of Lands, Natural Resources and Environmental and Protection, 2015; Galatowitsch, 2018). While natural phenomena such as storms, landslides, earthquakes, wave erosion, and water level rise can all result in changes to wetland regimes, human-caused changes to wetlands are far more prevalent (Galatowitsch, 2018). The land use/land cover (LULC) changes include the establishment of human settlements, the expansion of agricultural activities, and various other human-related actions (Winton *et al.*, 2021). These human-induced LULC changes can have devastating impacts on wetland biodiversity, hydrological regimes, and the essential services that wetlands provide to both human communities and the natural environment (Galatowitsch, 2018; Winton *et al.*, 2021).

The conservation of aquatic ecosystems and the supply of clean drinking water for human populations depend on the quality of the water. Water quality is crucial for maintaining healthy aquatic ecosystems and ensuring that water is safe for human consumption. Poor water quality can harm aquatic life and pose health risks to humans, making it an essential component of environmental conservation efforts (Gholizadeh *et al*., 2016). In order to ensure that the water is suitable for human consumption and aquatic life, water quality parameters must be monitored (Gholizadeh *et al*., 2016; Smith *et al.*, 2007). Water quality can be assessed through three primary types of metrics: physical, chemical, and biological. Physical factors include temperature, turbidity, color, taste, odor, electrical conductivity, salinity, and total dissolved solids. Chemical parameters encompass dissolved oxygen, total hardness, pH, acidity, alkalinity, sodium, chlorine, and others. Biological parameters comprise viruses, bacteria, algae, and faecal coliforms. These factors collectively determine the presence of pollutants, pathogens, and other microorganisms, as well as the overall chemical composition of the water, providing a comprehensive understanding of water quality (Pooja, 2018; Soeprobowati *et al.*, 2021).

Water quality is influenced by both point and nonpoint sources of pollution. Point sources include sewage discharge and industrial effluents, while nonpoint sources encompass atmospheric deposition, hydrological factors leading to runoff, and changes in LULC (Anh *et al*., 2023). These

sources of pollution, whether localized or diffuse, directly impact the biological, chemical, and physical properties of water bodies. This, in turn, affects the overall health of aquatic ecosystems and the availability of clean drinking water for humans and other organisms (Khatri and Tyagi, 2015; Anh *et al.*, 2023). Amidst the increasing urban and industrial development, understanding the close linkage between LULC and water quality becomes critical as water quality is crucial for various aspects of life, as it directly impacts human health, ecosystems, and economic activities like agriculture, tourism, and industry (Namugize *et al.*, 2018). Poor water quality can lead to severe health issues and can have detrimental effects on aquatic ecosystems as human activities and the natural environment interact in a dynamic manner, leading to significant implications for water quality (Khatri and Tyagi, 2015), making it essential to monitor and maintain high water quality standards (United Nations Environment Programme, 2008; Russ *et al.*, 2022). Therefore, this research aims to improve the understanding and predictive capabilities regarding how changes in LULC affect surface water quality in the Bangweulu Sub-Catchment. This goal is particularly significant given the ecological and socio-economic importance of the region, which is characterised by rich biodiversity and critical water resources.

1.2 Problem Statement

According to a comprehensive analysis of 189 research papers investigating wetland changes, between 64% and 71% of the world's wetlands have been lost since the early 1900s due to human activities, particularly changes in LULC patterns (Davidson, 2014). Projections indicate that wetland loss will continue at a similar rate going forward (Davidson, 2014; WWF, 2020). Zambia's wetlands, like the Bangweulu (meaning "water meets the sky") Wetland System (BWS), is one of the country's most highly prioritised sub-catchments for protection due to its ecological and economic importance (Lehner *et al.*, 2021). However, there is limited literature on LULC changes in the BWS. Accurately understanding LULC dynamics is complicated by the numerous classification techniques available, each with their own strengths and weaknesses. Non-parametric methods like Machine Learning (ML) offer greater accuracy, but different ML models have distinct advantages and disadvantages (Polikar, 2006; Rokach, 2010; Maxwell., 2018; Talukdar *et al.*, 2020). Remote sensing plays a crucial role in monitoring wetland water quality, but assessing several parameters in the major BWS lakes poses significant challenges, resulting in gaps and limitations. The Water Quality Index (WQI), which integrates various water quality parameters, has not been thoroughly evaluated for spatial application (Chidiac *et al.*, 2023). Traditional methods for assessing LULC's influence on surface water quality have limitations due to restricted spatial data coverage. Advancements in remote sensing and GIS technology help bridge this gap, but there is inadequate literature on the influence of LULC on WQI and limitations in the available methods (Cheng *et al.*, 2022). The widely used methods to

determine the influence of LULC on water quality share a common limitation of not considering the spatial distribution and proximity of LULC pollution sources to the point of interest (Liberoff *et al.*, 2019).

1.3 Research Objectives

1.3.1 Main Objective

To enhance the modeling of land-use/land-cover change and investigate its impact on surface water quality in the Bangweulu sub-catchment, Zambia.

1.3.2 Specific Objectives

i. To ensemble a superiorly hybrid machine learning model for enhanced accuracy of modelling LULC changes.
ii. To rapidly evaluate the water quality variability.
iii. To investigate the influence of LULC on water quality.

1.4 Research Questions

i. What combination of machine learning algorithms yields the highest accuracy in modelling LULC changes in wetland areas?
ii. How accurately can water quality parameters be estimated using empirical formula and remotely sensed water quality indices?
iii. How can multiple water quality parameter maps be integrated into a single Water Quality Index map?
iv. What is the relationship between different LULC types and water quality?

1.5 Significance of the Study

This study introduces an advanced ensemble hybrid machine learning model that significantly improves the accuracy of modeling LULC changes in wetland ecosystems. Specifically, the research developed a more precise LULC Quad hybrid map by integrating multiple input maps into a unified composite representation. Additionally, the study introduces an integrated approach for a comprehensive and rapid assessment of open water bodies, along with effective methods for exploring the relationship between various LULC types and water quality. This research output can support informed decision-making in wetland management and conservation worldwide, particularly in developing nations. It aims to benefit both current and future generations, aligning with the Zambian government's water policy of developing an effective national wetlands' conservation strategy. Furthermore, it supports sustainable wetland ecosystem management and is in line with the United Nations Sustainable Development Goals (SDGs) 2, 6,13, 14, and 15.

1.6 Conceptual Framework

A conceptual framework is a crucial component of research that provides a visual representation of the relationships between the key variables in a study. It outlines the expected connections and interactions between these variables, serving as a guiding roadmap for the research process (Jabareen, 2009). By illustrating the anticipated relationships between the central concepts, the conceptual framework helps bring together the various contexts and frames of a study in an explicit manner (Leshem and Trafford, 2007; Jozkowski, 2017). This conceptual framework helps to communicate the focus of the investigation and justify the importance of the research design. Moreover, this framework aids in the identification of gaps in the existing literature (Leshem and Trafford, 2007; Jozkowski, 2017). In this research, the remote sensing technology provided the satellite data, including Landsat and Sentinel images, which were processed using indices and empirical equations to estimate water quality parameters and calculate the water quality index. Field data was used to validate the satellite data. Additionally, the satellite data was also processed through a hybrid machine learning model to estimate LULC. The LULC data was used to evaluate its impact on both the water quality index and individual water quality parameters. Figure 1 presents a conceptual diagram that illustrates the key concepts and their expected relationships within the scope of this research.

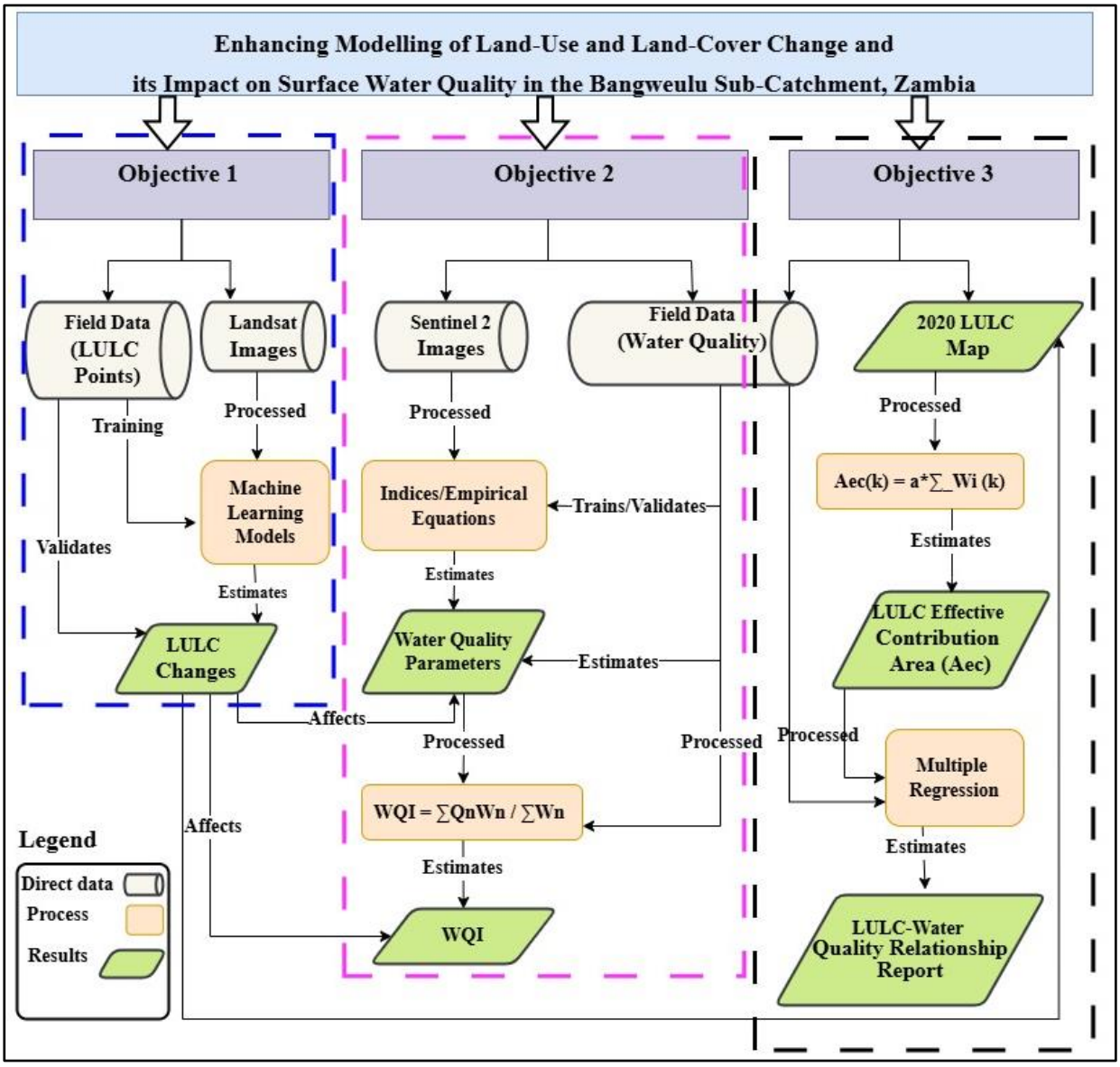


Figure 1: A conceptual diagram that illustrates the key concepts, the expected relationships, methods and results within the scope of this research.

1.7 Ethical Considerations

The University of Zambia Natural and Applied Sciences Research Ethics Committee (NASREC) approved this study, signifying that it met the ethical standards required for research involving human participants. The approval from NASREC not only legitimised the study but also aligned with international standards for ethical research (Appendix I).

1.8 Thesis Outline

Chapter 1 provides the background of the study, outlining the problem statement, main objective, specific objectives, research questions, significance of the study, conceptual framework, and ethical considerations. Chapter 2 presents a comprehensive literature review covering topics such as anthropogenic factors influencing land-use and land cover change, analysis of land-use and

land cover change, various machine learning algorithms, hybrid machine learning models, applications of machine learning in Africa, earth observation data for environmental monitoring, characteristics of Landsat and Sentinel sensors, water quality, and the influence of land-use and land cover (LULC) on water quality. Chapter 3 describes the study area, focusing on the physical and socio-economic characteristics of the Bangweulu wetland. In Chapter 4, the methodology is detailed, including the processes of data collection and analysis. Chapter 5 presents the results aligned with the study's objectives, while Chapter 6 discusses these results in depth. Finally, Chapter 7 concludes the thesis with a summary of findings, limitations, uncertainties, and recommendations. The report concludes with a reference section and appendices.

CHAPTER TWO: LITERATURE REVIEW

This section provides a comprehensive literature review, beginning with an examination of anthropogenic factors contributing to LULC change. It then explores on LULC change analysis, including various machine learning algorithms and their applications in Africa. Additionally, it discusses the use of Earth observation data for environmental monitoring, the characteristics of Landsat and Sentinel-2 sensors, and the impact of LULC on water quality. The section concludes with a summary of key research gaps identified in the literature that this study aims to address as well as the theoretical framework behind this study.

2.1 Anthropogenic Induced Factors of Land-use and Land Cover Change

Land-cover signifies the natural conditions of the land, including vegetation, man-made infrastructure (e.g., roads, bridges, or buildings), water, bare soil, and any other surfaces. In contrast, Land-use aligns with the purpose the land serves like recreation, wildlife habitat, or farming (Nedd *et al.*, 2021). LULC change is one of the pillars of human civilization as it involves the conversion of land to cultivate crops, raise animals, obtain timber, and build cities (Zorrilla-Miras *et al.*, 2014; Y. Zhang *et al.*, 2020). While LULC change provides these fundamental ecosystem commodities, it also has an impact on a variety of other ecosystem services that are critical to human survival like water quality. A critical understanding of water quality response to LULC change is required to balance the inherent trade-offs between meeting immediate human needs and maintaining other ecosystem functions. The impacts on the water quality vary depending on the type of LULC change and the biological context, and they have long-term local and global implications (Zorrilla-Miras et al., 2014; Y. Zhang et al., 2020).

The majority of Zambians, particularly those living in rural regions, heavily rely on the utilization of wetland natural resources for their livelihood due to rapid population growth and inadequate household income (Ministry of Lands Natural Resources and Environmental Protection, 2015). Phethi and Gumbo (2019) also observed that poverty and population increase drive wetland mismanagement in rural regions. The wetlands are predominantly used for crop cultivation, animal grazing, and fishing, leading to the deterioration of these ecosystems. Therefore, wetlands are among the ecosystems that have been under threat as a result of LULC changes (Foley *et al.*, 2005; Gagné and Fahrig, 2007). The Bangweulu wetlands have a number of islands that are inhabited and farmed. The island's vegetation has been overexploited and severely degraded into grasslands as a result of agriculture (primarily the Chitemene cultivation-slash and burn system), settlements, livestock grazing, and fires (Zambia Wildlife Authority, 2006). Human activity generally tends to leave biogeochemical impacts on its waterways as fertilizers and pesticides

leach from agricultural lands, and cultivation increases erosion and sedimentation, which lead to degradation of water quality through water eutrophication (Winton et al., 2021).

Banda et al. (2023) reported that LULC changes are among the many drivers that profoundly impact ecosystem services by altering the natural functions of wetlands. Davidson (2014) also reported that 64 –71% of wetlands have been lost worldwide since 1900 due to anthropogenic factors. Zhang et al. (2021) added that the anthropogenic factors such as urban growth, has led to wetlands being drained, filled, or altered, resulting in the loss of valuable wildlife habitats. Furthermore, 80% of global wastewater enters wetlands, posing health risks (FAO, 2017; Xu *et al.*, 2019). Xu *et al.* (2019) also added that these factors actually threaten the health and survival of the major wetlands. Therefore, Maitima *et al.* (2010) concluded that LULC changes can have detrimental effects such as modifying the wetlands' ability to control floods, disrupting sediment and nutrient transport, compromising aquatic biodiversity, obstructing the migration routes of aquatic organisms, increasing soil salinity, making wetlands more susceptible to erosion as well as disturbing the connectivity of river systems.

Zambia has eight wetlands designated as wetlands of international importance as earlier alluded to, which are spread throughout seven river basins (Ngoma *et al.*, 2017; Yang *et al.*, 2021). These wetlands, especially the Bangweulu (meaning “water meets the sky”) Wetland System (BWS), are critical for biodiversity and environmental services. The BWS is one of Zambia's most highly prioritised sub-catchments for protection (Lehner *et al.*, 2021). However, the BWS faces the threats of LULC changes such as excessive exploitation of ecosystem services by local populations, unsustainable development and fishing practices, deforestation, and agricultural expansion (Kamweneshe, Beilfuss and Morrison, 2003; Zambia Wildlife Authority, 2006). Despite these threats, literature is inadequate on the LULC changes related to the BWS and its surrounding areas. The surrounding areas consist of regions where streams flow into the BWS (Kamweneshe, Beilfuss and Morrison, 2003; Zambia Wildlife Authority, 2006). This knowledge gap complicates understanding and effectively managing the dynamic nature of the BWS.

2.2 Land-Use and Land Cover Change Analysis

Timely and accurate detection of changes in the BWS surface features is critical for wetland conservation, sustainable development, and water resource management as well as understanding the interactions between human and natural phenomena (Lu *et al.*, 2004; Zhu *et al.*, 2019). Recent advances in remote sensing tools and techniques enable researchers to detect and monitor such changes at a large scale, as remote sensing data like Landsat data are used as the primary source for change detection (Hemati *et al.*, 2021). The Landsat data are crucial for

environmental and ecological monitoring as they provide early baseline data for change detection in areas with valuable ecosystems (Wulder *et al.*, 2019; Hemati *et al.*, 2021).

It should be noted that for several decades, the statistical approach, specifically parametric methods, has served as the conventional method for fitting models in LULC classification. This approach operates under the assumption that the data is generated using a stochastic data model (Loussaief and Abdelkrim, 2017). The challenge with this approach is that they work with a pattern that can be statistically well analysed but can not be predicted precisely. As a result, statistical data model conclusions are about the model's mechanism rather than nature's mechanism; if the model is a poor emulator of nature, the conclusion may be incorrect (Breiman, 2001b; Loussaief and Abdelkrim, 2017). Furthermore, statistical models in spatial LULC analysis are problematic because they assume that data is statistically independent and uniformly distributed. However, spatial LULC data are highly heterogeneous and tend to be dependent, a phenomenon known as spatial autocorrelation (Charif *et al.*, 2012). To address this challenge, machine learning algorithms like non-parametric methods are considered.

2.3 Machine Learning Algorithms

Machine Learning (ML) is a field of artificial intelligence (AI) and computer science that is based on the idea that systems can learn from data and algorithms modelling culture to mimic how humans learn, identify patterns, and make decisions (Chollet, 2017). The use of ML models in this research was primarily motivated by the models' universal approximation capabilities and high performance in a wide range of scientific fields (Charif *et al.*, 2012). ML employs algorithmic models that treat the data mechanism as unknown, complex, and highly heterogeneous. In addition, algorithmic modelling can be used on large complex data sets as well as on smaller data sets as a more accurate and informative alternative to statistical data modelling (Breiman, 2001b; Loussaief and Abdelkrim, 2017).

ML algorithms can be used to predict how ecosystems respond to changes in environmental variables by detecting nonlinear empirical relationships between variables based on a set of representative data sets (Sadiq, Rodriguez and Mian, 2019; Antunes *et al.*, 2021). Hence, ML algorithms prove to be valuable tools for modelling complicated ecosystems like wetlands. Their advantages encompass operational simplicity, reliability in handling nonlinearity, robustness, and the ability to handle noisy data (Kang, Rule and Noble, 2012; Pinto, Antunes and Roca, 2021). There are several classification algorithms used in ML which include: (1) Artificial Neural Networks (ANNs), (2) Naïve Bayes (NB), (3) Support Vector Machines (SVM), (4) Decision

Tree (DT), (5) Random Forest (RF), and (6) K-Nearest Neighbour (KNN) algorithms (Wang *et al.*, 2006; Zou, Han and So, 2008; Qiu, Jiang and Li, 2015; Chen *et al.*, 2020).

2.3.1 Types of Machine Learning Algorithms

Artificial Neural Network Algorithm

Artificial neural networks (ANNs) are machine learning nonlinear modelling techniques based on the function of the human nervous system which allows learning by example from sample data that describes a naturally occurring phenomenon (Zou, Han and So, 2008). ANN is made up of three layers: (1) input layer, (2) output layer, and (3) one or more hidden layers (Figure 2). By firing activation functions, the input layer nodes pass information to the hidden layer nodes, and hidden layer nodes either fire or remain dormant depending on the evidence presented. The evidence is weighted by the hidden layers, and when the value of the set of nodes in the hidden layer reaches a certain threshold, it is passed on to one or more nodes in the output layer (Ahmed, Dey and Sarma, 2011; Sadiq, Rodriguez and Mian, 2019; Shafizadeh-Moghadam, 2019).

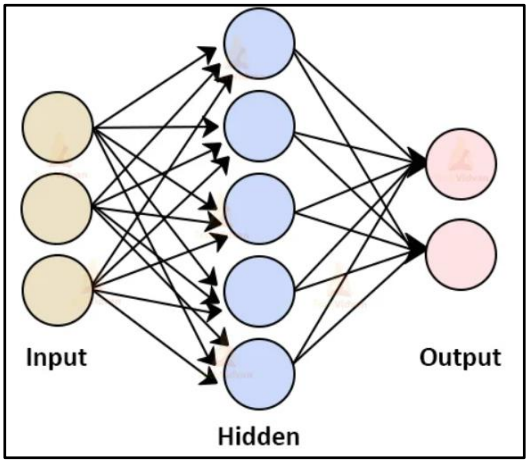


Figure 2: The Concepts of Artificial Neural Network Architecture

Adapted from Senthilkumar (2010)

ANNs have the following advantages: they are flexible and can be used for both regression and classification problems; they are good to model nonlinear data with a large number of inputs; they are reliable in an approach of tasks involving many features; and the predictions are fairly fast once the sample has been trained. However, neural networks face the following difficulties: They are black boxes, so you don't know how much each independent variable influences the dependent variables; training samples with traditional computers is computationally expensive and time-consuming; they rely heavily on training data because they work best with more data points. This creates the issue of over-fitting and generalization (Tu, 1996; Dumitru and Maria, 2013).

Naive Bayes

The classification method known as Naive Bayes (NB) is based on Bayes' theorem (figure 3) and makes the assumption that all of the features that predict the target value are unrelated to one another. While the NB classifier calculates the likelihood of each class before selecting the one with the highest probability, it ignores the fact that features in real-world data depend on one another in choosing the target. Despite the fact that the independence assumption is never true for data from the real world, it frequently works effectively in practice. Thus, it is referred to as "Naive" (Wang *et al.*, 2006; Qiu, Jiang and Li, 2015; Chen *et al.*, 2020).

The NB classifier has several strengths which include being simple to implement, conditional probabilities are simple to evaluate, Iterations are unnecessary because the probabilities can be calculated directly, if the conditional independence assumption is correct, the results could be spectacular, its robust to noise feature, requires a small amount of training data to estimate the classification parameters (Harrington, 2015; Kalcheva, Todorova and Marinova, 2020). It does, however, have the following drawbacks: The conditional Independence assumption does not always hold because the feature is dependent in most cases. Therefore, if the feature are highly correlated, it performs poorly (Kalcheva, Todorova and Marinova, 2020).

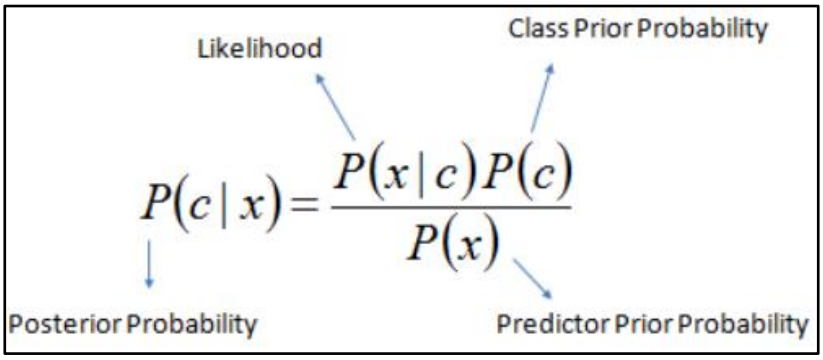


Figure 3: The Mathematical Expression of Bayes' Theorem

Adapted from Ballesteros-Pérez et al. (2018)

Support Vector Machine

Support Vector Machines (SVMs) are supervised learning models used to solve linear and nonlinear classification and regression analysis of complex data patterns. SVM can handle multiple continuous and categorical variables and can support both regression and classification tasks because it has three different types of kernel functions (linear, polynomial, radial basis function, and sigmoid function) (Roy, Kar and Das, 2015; Gholami and Fakhari, 2017). SVM works by creating a separating line (hyperplanes) in a multidimensional space to separate cases

with different class labels and selects the points (support vectors) from both classes that are closest to the line. The distance between the line and the support vectors is then calculated by the algorithm. This distance is known as the margin, and the goal is to maximize it (Figure 4). The optimal hyperplane is the one with the greatest margin. As a result, SVM attempts to create a decision boundary with the greatest possible separation between the two classes. Even when the number of dimensions exceeds the number of samples, SVMs are effective in high-dimensional spaces. They have a low memory footprint and are adaptable. However, if the number of features exceeds the number of samples, the method is likely to perform poorly (Chen *et al.*, 2010; Pisner and Schnyer, 2019). The SVM classifier has the following advantages and disadvantages: It works well when there is a clear dividing line between classes. It is more effective in high-dimensional spaces, and it is effective when the number of dimensions exceeds the number of samples. However, the SVM algorithm is not appropriate for large data sets. When the data set contains more noise, such as overlapping classes, it does not perform well. Furthermore, when the number of features for each data point exceeds the number of training data samples, the SVM performs poorly (Anguita *et al.*, 2010; Pereira and Joshi, 2014)

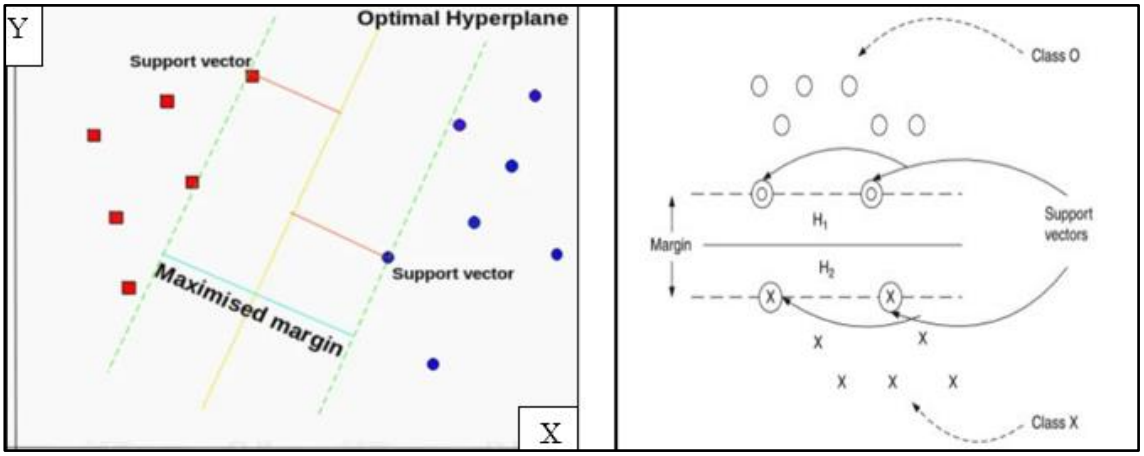


Figure 4: Support Vector Machine Architecture

Adapted from Chen *et al.* (2010)

Decision Tree

Decision Tree (DT) algorithm is a type of Supervised Machine Learning in which data is constantly split based on a specific parameter. Two entities can be used to explain the tree: decision nodes and leaves as illustrated in Figure 5. The decisions or outcomes are represented by the leaves, and the data is split at the decision nodes (Szczerbicki, 2001; Du and Sun, 2008). The decision tree classifier's main advantage is its ability to use different feature subsets and decision rules at different stages of classification. The first step in the DT classifier is determining the entropy of the database, which tells how uncertain the database is. The lower the value of

uncertainty, the better the classification results. The information gained from each feature is computed. This tells how much uncertainty has been reduced as a result of spitting the database. Finally, all of the information gained is calculated for all features, and the database with the highest information gain is split. The process is repeated until all nodes have been cleared (Szczerbicki, 2001; Kamble and Dale, 2022). The DT algorithm makes nonlinear data patterns very easy to understand and visualize. They are also very fast, especially when it comes to exploring data. Decision trees provide a high accuracy score even if the dataset is small. However, one of decision trees' constraints is that they are relatively unstable. A minor change in the data can cause a significant change in the structure of the decision tree, resulting in a different result than what users would receive in a typical event (Rokach and Maimon, 2006; Prajwala, 2015).

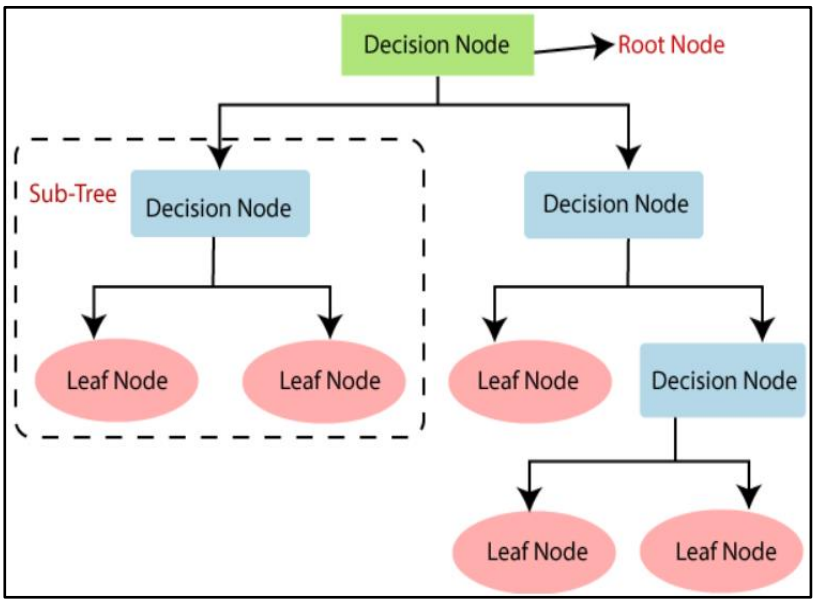


Image source: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

Figure 5: Decision Tree Architecture

Random Forest

The Random Forest Algorithm is a supervised machine learning algorithm composed of tree predictors in which each tree is dependent on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Figure 6). Random Forests can be used for categorical response variables (classification) or continuous response variables (regression). Likewise, predictor variables can be categorical or continuous; when using Random Forest for classification, each tree assigns a classification, and the forest selects the classification with the highest number of votes. When using Random Forest for regression, the forest chooses the average of all tree outputs (Cutler and Stevens, 2012).

Random Forest generates multiple decision trees, which are then combined to make a more accurate prediction. The Random Forest model is based on the idea that multiple uncorrelated

models (individual decision trees) perform much better as a group than they do individually as individual decision trees may make mistakes, but, the majority of the group will be correct, thereby moving the overall result in the right direction (Cutler and Stevens, 2012; Louppe, 2014).

Among the finest ML classifiers is RF. It manages large datasets with tens of thousands of variables. When a class is more infrequently observed in a data collection than other classes, it can automatically balance the data sets. It has a relatively low risk of overfitting and performs well with non-linear data (Rodriguez-Galiano *et al.*, 2012). However, RF often uses random grouping and accurately implements decision trees, but due to the extrapolation issues, all test data points outside of the training area are not included in the predictions. RF is also found to be biased when dealing with categorical variables and is not appropriate for linear methods with a large number of sparse features (Prajwala, 2015).

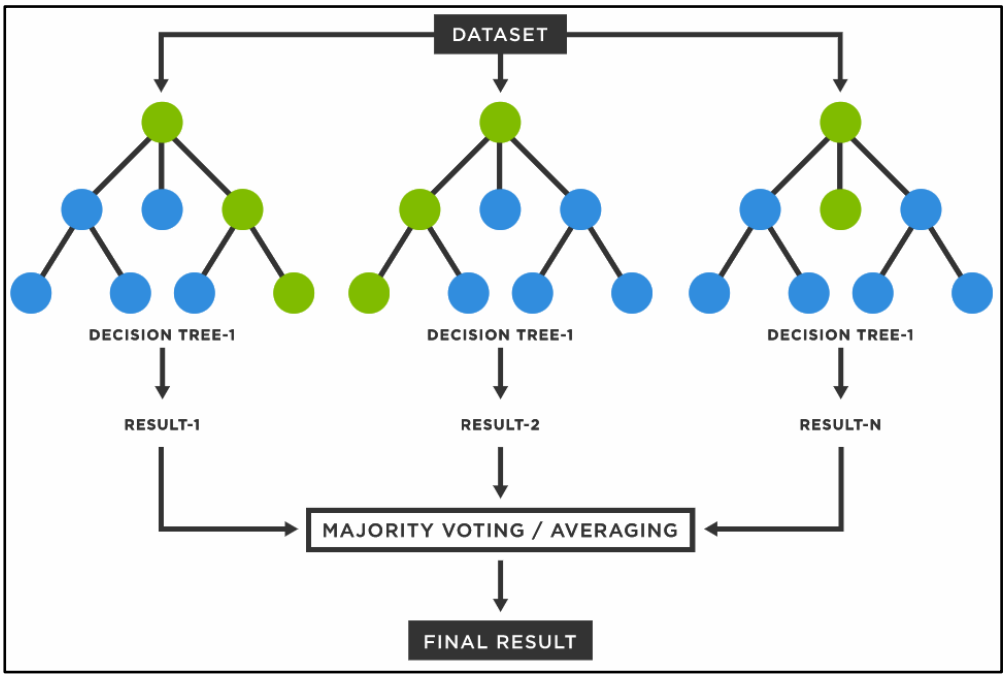


Figure 6: Random Forest Architecture

Image source: https://www.tibco.com/reference-center/what-is-a-random-forest

K – Nearest Neighbour

The k-Nearest Neighbours algorithm is a non-parametric, supervised learning classifier that uses proximity to make classifications or predictions about the grouping of an individual data point because it assumes that similar things exist in close proximity, and it is used to solve classification and regression problems (Khan, Ding and Perrizo, 2002; Cunningham and Delany, 2021). If there are two groups, A and B, KNN algorithm examines the states of data points nearby to determine whether a data point belongs in group A or group B. If the majority of data points are in group

B, the data point in question is almost certainly in group B, and vice versa (Figure 7). Algorithm works well when the data points are either well defined or non-linear (Cunningham and Delany, 2021).

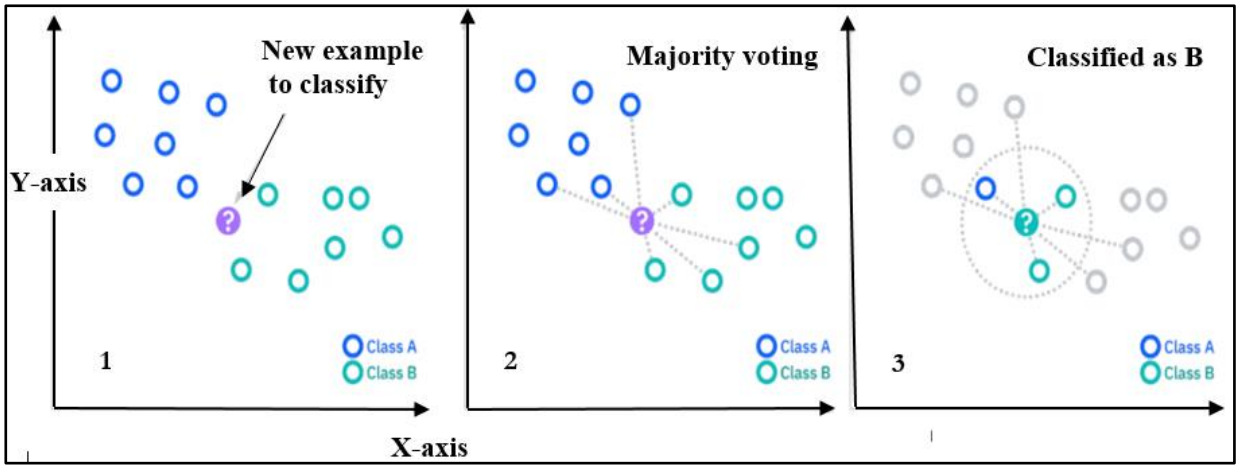


Figure 7: KNN Predictions Procedure

Boosting Algorithms

Boosting algorithms seek to improve prediction power by training a sequence of weak models (Figure 8), each compensating for the shortcomings of its predecessors because it assumes that a weak learner's combination of simple classifiers can outperform any of the simple classifiers alone. It assigns weights to individual tree outputs, then gives incorrect classifications from the first decision tree a higher weight and input to the next tree. The boosting method combines these weak rules into a single powerful prediction rule after several cycles (Ferreira and Figueiredo, 2012; Mayr *et al.*, 2014). The boosting algorithm's replica methods, such as decision trees and random forests, effectively increase its predictive power. Moreover, boosting is a strong technique that readily reduces over-fitting. However, as every classifier is required to correct the mistakes made by the forerunners, boosting has the drawback of being sensitive to outliers. The method's difficulties to scale up is another drawback. This is because it is challenging to simplify the procedure because every estimator rests its accuracy on the prior predictions (Schapire *et al.*, 1999; Ferreira and Figueiredo, 2012).

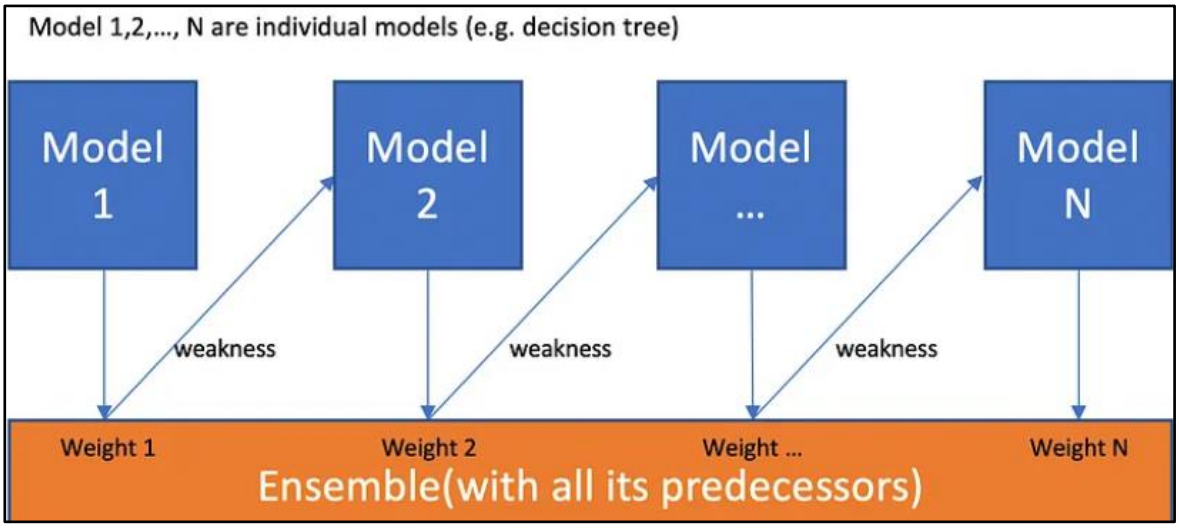


Figure 8: Boosting Algorithms Architecture

Image source (https://towardsdatascience.com/boosting-algorithms-explained-d38f56ef3f30)

Hybrid ML model

ML models, being diverse in nature, possess unique sets of strengths and weaknesses that set them apart from one another. These distinctions arise due to differences in algorithmic approaches, training data requirements, interpretability, generalisation capabilities, computational efficiency, methods, time and space, and robustness to noise (Cutler, Cutler and Stevens, 2012; Qiu, Jiang and Li, 2015; Chen *et al.*, 2020). Therefore, experimentation and evaluation of different ML models on the specific task are crucial to determine the best model that is capable of achieving reliable accuracy (Maxwell, Warner and Fang, 2018; Camargo *et al.*, 2019; Talukdar *et al.*, 2020). The multiple best models can then be integrated into a hybrid model which would allow for the creation of a more powerful and accurate model by leveraging the diverse strengths and weaknesses of different models (Polikar, 2006; Rokach, 2010).

2.4 Application of ML Models in Africa

The use of ML models in Africa has shown success in improving LULC classification performance (Boateng, Otoo and Abaye, 2020; Mahmoud *et al.*, 2023; Yuh *et al.*, 2023). However, these methods often require significant image pre-processing, especially with coarse-resolution images, to reduce uncertainties in LULC classifications. Furthermore, there has been limited application of these approaches to effective monitoring of changes in LULC within wetland areas across Africa (Yuh *et al.*, 2023), where coarse-resolution satellite images are frequently the only available option. Existing studies have generally relied on applying only a single method (Thamaga, Dube and Shoko, 2022; Gxokwe, Dube and

Mazvimavi, 2023; Yuh *et al.*, 2023), which can increase classification uncertainties relative to the use of an ensemble (hybrid) approach.

2.5 Earth Observation data for environmental monitoring

The number of active and passive remote sensing satellites has increased due to advancements in space-borne instrument technology and the growing importance of earth observation (EO) satellites for environmental monitoring. The major advantage of using remote sensing is that it provides a synoptic view, large spatial coverage, and frequent observations which results in large data volumes and multiple data sets with varying spatial and temporal resolutions (Belward and Skøien, 2015; Gómez, White and Wulder, 2016; Strauss, 2017). However, the majority of these satellite sensor programs are costly, lack radiometric and spectral band passes consistency, orbital stability, and necessary reference historical data which are needed for accurate Land-use change detection (Wulder *et al.*, 2018). The Landsat program addresses these issues by providing radiometric consistency, a large historical record, and continuous data collection at the national and global levels (Wulder *et al.*, 2018). The Landsat program is one of the few instruments capable of serving as a reference sensor for other optical satellite programs, providing accurate data to policymakers and governments who increasingly use LULC change detection techniques for national reporting and policy development (Wulder *et al.*, 2019). Furthermore, the satellites have a very good ground resolution and spectral bands for tracking Land-use change caused by natural and anthropogenic factors (United States Geological Survey, 2022).

The Landsat program (Figure 9) began in 1972, providing an early baseline for change detection in locations that are unique among satellite programs archive (Belward and Skøien, 2015; Wulder *et al.*, 2019; Hemati *et al.*, 2021). Nearly five (5) decades, the Landsat satellites have been observing the Earth's environment, providing detailed images of most of our planet's valuable ecosystems such as forests, wetlands, coastlines, glaciers, volcanoes, and oceans (*Environmental Watch with Landsat satellites*, 2014). The United States Geological Survey later changed the Landsat data policy in 2008, resulting in free and open access to the global Landsat archive (Woodcock *et al.*, 2008). Despite the current growing number of EO satellites, including high-resolution commercial satellites, the Landsat program continues to play an important role in optical imagery applications, serving as an indispensable reference point. This is because there is no other science-grade, long-term uninterrupted EO program that offers a free data policy, global coverage, radiometric uniformity between sensors, and acceptable spatial and temporal resolutions for the environment and ecological monitoring

(Wulder *et al.*, 2019; Hemati *et al.*, 2021). The Landsat mission's timeline and the previously launched satellites are summarised in Figure 9.

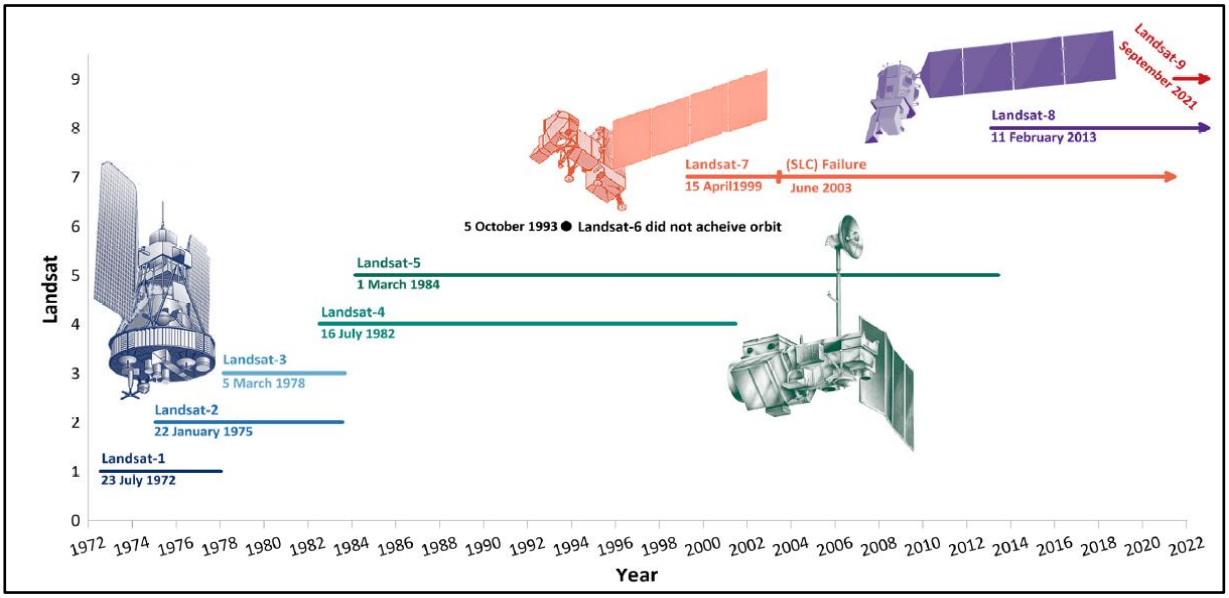


Figure 9: Landsat Mission's Timeline from 1972 to 2022 Source: (Hemati *et al.*, 2021)

Remote sensing technology for monitoring water quality parameters in open water bodies offers numerous advantages, including improved frequency of monitoring, real-time data acquisition, and synoptic view of spatial and temporal coverage of the water body that is unattainable by conventional methods, making it an ideal and cost-effective way for integrated water quality assessment of large water bodies (Gholizadeh *et al*., 2016). The choice of remote sensing platform and sensor plays a crucial role in water quality monitoring. Satellite-based sensors, such as those on board the Sentinel mission, are used to retrieve water quality parameters and provide spatial dataset with global coverage, making it valuable for long-term water quality monitoring (Bid and Siddique, 2019; Cheng *et al.*, 2022). The sentinel 2 which was initially designed for terrestrial applications, the significantly improved spatial footprint and the additional bands in the short-wave infrared region of Sentinel-2 satellite missions offer an optimal opportunity to detect the dynamic spatial heterogeneity that characterizes terrestrial-aquatic environments and their interfaces at local, regional, and global scales (Page *et al*., 2019; Vanhellemont, 2019). Moreover, the Sentinel-2 has an exceptional high spatial resolution, which facilitates detailed mapping of water bodies and their surroundings. Its multispectral capabilities, which allow data collection across various spectral bands, enable the estimation of different water quality parameters. Additionally, its real-time data availability is crucial for timely monitoring and forecasting of water quality changes. Furthermore, Sentinel-2 data is cost-effective due to its free accessibility and large-area coverage (Casal, 2022; Quang *et al.*, 2023).

2.6 Characteristics of Landsat and Sentinel Sensors

Landsat 5 carried the Thematic Mapper (TM) (Table 1) and Multispectral Scanner (MSS) (Table 2) sensors. It orbited the Earth in a sun-synchronous, close-to-polar orbit at 705 kilometers (438 miles) (98.2 degrees inclination). In addition, it had a 16-day repeat cycle, a 9:45 a.m. (+/- 15 minutes) equatorial crossing time, and it orbited the Earth every 99 minutes with 14 orbits per day. Furthermore, it acquired data on the Worldwide Reference System-2 (WRS-2) path/row system and its swath overlap (or side lap) varied from 7% at the equator to a maximum of 85% at extreme latitudes (United States Geological Survey, 2013; Zhou *et al.*, 2006).

Table 1: Landsat 5 TM Band Characteristics.

Landsat 5 TM		
Bands	**Wavelength (micrometers)**	**Resolution (meters)**
Band 1 - Blue	0.45-0.52	30
Band 2 - Green	0.52-0.60	30
Band 3 - Red	0.63-0.69	30
Band 4 - Near Infrared (NIR)	0.76-0.90	30
Band 5 - Shortwave Infrared (SWIR) 1	1.55-1.75	30
Band 6 - Thermal	10.40-12.50	120* (30)
Band 7 - Shortwave Infrared (SWIR) 2	2.08-2.35	30

(United States Geological Survey, 2013)

Table 2: Landsat 5 MSS Band Characteristics.

Land 5 MSS		
Bands	**Wavelength (Micrometres)**	**Resolution (meters)**
Band 1 - Visible Green	0.5 – 0.6	60
Band 2 - Visible Red	0.6 – 0.7	60
Band 3 - NIR	0.7 – 0.8	60
Band 4 - NIR	0.8 – 1.1	60

(United States Geological Survey, 2013)

The Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) instruments are aboard the Landsat 8 satellite (Table 3). The OLI measures in the visible, near infrared, and shortwave infrared (VNIR, NIR, and SWIR) regions of the spectrum. The TIRS measures the ground surface temperature in two thermal bands. It orbits the earth in a close-to-polar, sun-synchronous orbit (98.2 degrees inclination) at a height of 705 kilometers. In every 99 minutes, one Earth orbit is completed and crosses the equatorial at 10:00 a.m. +/- 15 minutes, 16-day repeat cycle thereby acquiring 740 scenes every day on the WRS-2 path/row system, with a swath overlap (or side lap) ranging from 7% at the equator to a maximum of nearly 85% at extreme latitudes (United States Geological Survey, 2013)

Table 3: Landsat 8 Band Characteristics.

Landsat 8 OLI and TIRS		
Bands	**Wavelength (micrometers)**	**Resolution (meters)**
Band 1 - Ultra Blue (coastal/aerosol)	0.435 - 0.451	30
Band 2 – Blue	0.452 - 0.512	30
Band 3 - Green	0.533 - 0.590	30
Band 4 - Red	0.636 - 0.673	30
Band 5 - Near Infrared (NIR)	0.851 - 0.879	30
Band 6 - Shortwave Infrared (SWIR) 1	1.566 - 1.651	30
Band 7 - Shortwave Infrared (SWIR) 2	2.107 - 2.294	30
Band 8 - Panchromatic	0.503 - 0.676	15
Band 9 - Cirrus	1.363 - 1.384	30
Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30)

(United States Geological Survey, 2013)

The Sentinel-2 mission consists of two identical satellites, Sentinel-2A and Sentinel-2B, each carrying a Multispectral Instrument (MSI) with the following key characteristics: 13 spectral bands in the visible, near infrared (VNIR), and short-wave infrared (SWIR) spectral ranges; spatial resolution of 10 m (4 bands), 20 m (6 bands), and 60 m (3 bands); 290 km swath width; 12-bit radiometric resolution (0-4095 brightness levels); instantaneous field of view of about 21° by 3.5^{o} (Kaplan and Avdan, 2017; Romero, Marcello and Vilaplana, 2020). The two satellites operate in the same sun-synchronous orbit at an altitude of 786 km, phased 180° apart, allowing for a 5-day revisit time at the equator and 2-3 days at mid-latitudes. This enables systematic global coverage of land surfaces from 56°S to 84°N (Kaplan and Avdan, 2017; Romero, Marcello and Vilaplana, 2020). Table 4 summarises sentinel 2 bands and their description and spatial resolution.

Table 4: Sentinel 2 Bands, Spatial Resolution, Central Wavelength and Description

Sentinel-2 Bands	Resolution (m)	Central Wavelength (nm)	Description
Band 1	60	443	Ultra Blue (Coastal and Aerosol)
Band 2	10	490	Blue
Band 3	10	560	Green
Band 4	10	665	Red
Band 5	20	705	Vegetation Red Edge
Band 6	20	740	Vegetation Red Edge
Band 7	20	783	Vegetation Red Edge
Band 8	10	842	Near Infrared
Band 8a	20	865	Vegetation Red Edge
Band 9	60	940	Short Wave Infrared (Water vapour)
Band 10	60	1375	Short Wave Infrared (Cirrus)
Band 11	20	1610	Short Wave Infrared (SWIR 1)
Band 12	20	2190	Short Wave Infrared (SWIR 2)

(Kaplan and Avdan, 2017)

2.7 Water Quality

The conservation of aquatic ecosystems and the supply of clean drinking water for human populations depend on the quality of the water, which is an essential component of the environment (Gholizadeh *et al*., 2016). In order to ensure that the water is suitable for human consumption and aquatic life, water quality parameters must be monitored (Smith *et al.*, 2007). There are three primary types of water quality metrics: physical, chemical, and biological. Physical factors like temperature, turbidity, colour, taste, odour, electrical conductivity, salinity, and total dissolved solids. Chemical parameters include dissolved oxygen, total hardness, pH, acidity, alkalinity, sodium, and chlorine among others. Biological parameters comprise viruses, bacteria, and algae as well as faecal coliforms. The presence of pollutants, pathogens, and other microorganisms in the water, as well as the chemical makeup of the water, are all determined by these factors (Pooja, 2018; Soeprobowati et al., 2021).

The Water Quality Index (WQI) is a mathematical model that integrates water quality parameters into a single integer value, depicting the overall health status of a water body. The WQI is used to describe the water quality based on the physical, chemical, and biological factors of water quality, combined into a score varying from 0 to 100 (Marselina, Wibowo and Mushfiroh, 2022; Chidiac *et al.*, 2023). Despite the absence of a universally accepted WQI, one of the commonly used methods was developed by Brown *et al.*, (1970). This equation calculates the index based on particular water quality parameters and different

research studies adopt three main systems for parameter selection (Nguyen Van *et al.*, 2022). The three main systems include (1) a fixed system that uses a predefined set of nine basic water quality parameters (Swamee and Tyagi, 2007). However, the fixed system lacks the flexibility to include additional parameters specific to a location (Nguyen Van *et al.*, 2022), (2) Open system where, unlike the fixed system way of parameter selection, the open system has no particular rule in its selection of parameters. It enables the researchers to consider any water quality parameters of their choice that apply to the study, (3) Mixed system which is a hybrid design that combines both the nine basic water quality parameters and additional parameters, thereby embracing the mix of rigidity and flexibility of water quality parameters selection (Dojlido *et al*, 1994; Canadian Council of Ministers of the Environment., 2001; Nguyen Van *et al.*, 2022).

Although WQI integrates various water quality parameters for easier interpretation of water quality, its spatial application has not yet been thoroughly evaluated, which is a major knowledge gap that impacts the accuracy of the model due to variations in values across different locations that are influenced by different geographical, climatic, and human factors (Chidiac *et al.*, 2023). This limitation results in inaccurate water quality assessments and hinders the identification of specific areas with high levels of contamination (Swamee and Tyagi, 2007; Nguyen Van *et al.*, 2022; Chidiac *et al.*, 2023). Therefore, understanding the spatial distribution of water quality factors is essential for effective water quality management. To address this challenge, the use of remote sensing to analyse and visualise the spatial distribution of WQI values allows for a more detailed understanding of the spatial variation in water quality and helps identify areas that require further monitoring and intervention as recommended by Chidiac *et al.* (2023) and Ramadas and Samantaray (2018). Figure 10 shows the priciples of open water bodies remote sensing.

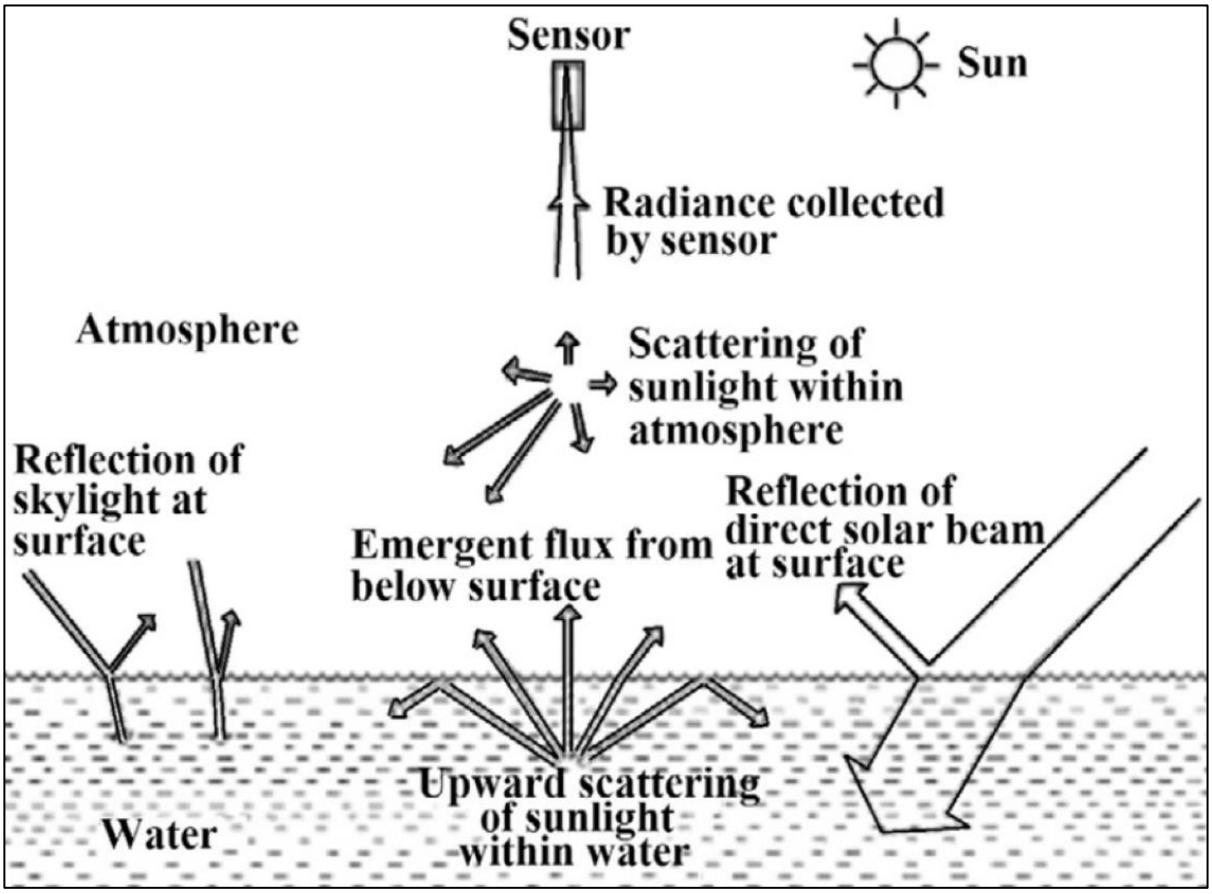


Figure 10: Principles of open water remote sensing (Yang et al., 2022)

There are several empirical equations where remote sensing techniques are used to retrieve the relationship between empirical equations and the water quality parameters. The Normalized Difference Water Index (NDWI) is used to highlight open water features in satellite imagery by allowing a water body to stand out against the soil and vegetation (McFeeters, 1996; Xu, 2006a). It is used to detect and monitor minor changes in open water bodies in a satellite image by using the near-infrared (NIR) and green spectral bands. The index's disadvantage is that it is sensitive to built-up structures and sand, which leads to overestimation of the water bodies (Bahrawi and Elhag, 2019; Xu, 2006). To address the limitations of NDWI, Xu (2006) developed a better water index than NDWI, the Modified Normalized Difference Water Index (MNDWI), which employs the Shortwave Infrared band rather than the NIR band. MNDWI improves open water feature extraction while effectively restricting and even eliminating built-up land noise, vegetation noise, and soil noise bodies (Bahrawi and Elhag, 2019; Xu, 2006).

The use of spectral reflectance from multiple bands such as band ratios rather than single bands has been observed to enhance the signal-to-noise ratio and decrease the impact of the atmosphere on water quality parameter estimation (Gholizadeh *et al*., 2016; Chawla, Karthikeyan and Mishra,

2020; Adjovu *et al.*, 2023). The Normalized Difference Salinity Index (NDSI) is a remote sensing index that is widely used for estimating the salinity (concentration of dissolved salts) of the soil (Allbed, Kumar and Aldakheel, 2014; Ennaji *et al.*, 2018; Gerardo and de Lima, 2022) and to a lesser extent salinity of the water bodies (de Baar, van Heuven and Middag, 2018; Al-Jabri *et al.*, 2023). Sodium and chloride are the major salt ions that are commonly found in water and are related to salinity (Ullman, 2013; de Baar, van Heuven and Middag, 2018). As earlier alluded to, the NDSI has been mostly used for estimating soil salinity, however, it has the potential in estimating the water salinity (Al-Jabri *et al.*, 2023) which can then be used to estimate the concentration of sodium and chlorides in water. Furthermore, the index can also be used to determine water electrical conductivity (EC) and total dissolved solids (TDS). EC measures water's ability to conduct an electrical current and correlates to the concentration of dissolved ions (sodium and chlorides) in the water. TDS is a measure of the total amount of dissolved solids in water, hence, the dissolved ions affect both EC and TDS, therefore, EC and TDS are correlated (Rusydi, 2018). The Normalised Difference Turbidity Index (NDTI) is a numerical indicator that uses remote sensing data to evaluate the turbidity or clarity of water. Turbidity simply means the presence of suspended particles in water, such as sediments, algae, or other substances that scatter and absorb light, reducing water clarity (Bid and Siddique, 2019). Research has shown that the NDTI is effective in analysing turbidity levels in various water bodies as well as a useful tool for monitoring water quality conditions using remote sensing technology (Bid and Siddique, 2019; Powers et al., 2023).

Although remote sensing techniques play a crucial role in monitoring water quality, there are limitations in the parameters monitored as well as their integration, which leads to a limited understanding of overall water quality conditions (Gholizadeh *et al.*, 2016; Ramadas and Samantaray, 2018). Integration of a broader spectrum of water quality parameters for water quality accurate assessment enhances decision-making and effective management strategies in the conservation of water resources (Chidiac *et al.*, 2023; Syeed *et al.*, 2023). For example, Ahmed *et al.* (2023) and Wang *et al*. (2017) estimated several water quality parameters and WQI using remote sensing data. However, their studies only displayed the spatial distribution of localised WQI values from sampling points only.

2.8 Influence of LULC on Water Quality

Water quality is crucial for various aspects of life, as it directly impacts human health, ecosystems, and economic activities like agriculture, tourism, and industry. Poor water quality can lead to severe health issues and can have detrimental effects on aquatic ecosystems, making it essential to monitor and maintain high water quality standards (United Nations Environment

Programme, 2008; Russ *et al.*, 2022). Human activities and the natural environment interact in a dynamic manner, leading to significant implications for water quality (Khatri and Tyagi, 2015).

In both rural and urban regions, water quality is affected by both point and nonpoint sources. Point sources of pollution include sewage discharge and industrial discharge, whereas nonpoint sources include, atmospheric deposition, hydrological factors that lead to run-off as well as LULC changes (Anh *et al.*, 2023). The biological, chemical, and physical properties of water bodies are affected by these sources of pollution, which in turn affects ecosystem health and the availability of clean drinking water for humans and other organisms (Anh et al., 2023; Khatri & Tyagi, 2015). Amidst the increasing urban and industrial development, understanding the close linkage between LULC and water quality becomes critical (Namugize, Jewitt and Graham, 2018).

Surface water quality is one of the key measures of ecosystem integrity, human well-being, and sustainable development. It depicts the impacts of human activity through LULC changes, pollutant runoff, and hydrological interactions (Mello *et al.*, 2020). Usually, the conventional methods in assessing the influence of LULC changes on surface water quality tend to suffer from low spatial coverage data (Cheng *et al.*, 2022). Nevertheless, the progress in remote sensing technology as well as geographic information systems (GIS) overcome this gap by combining field data with satellite imagery (Cheng *et al.*, 2022).

The influence of LULC on surface water quality has been thoroughly studied in recent years, with a focus on the relationship between LULC and water quality indicators (Ahmad *et al.*, 2021; Wang *et al.*, 2021; Cheng *et al.*, 2022). However, there is inadequate literature on the assessment of the Influence of LULC on the overall water quality (Water Quality index). In addition, other major research gaps include (1) the need for more comprehensive studies focusing on identification of the utmost accurate key water quality indicators that are most vulnerable to changes in LULC, thereby providing a more targeted approach to monitoring and managing water quality (Gani *et al.*, 2023), (2) establishing the linkage between LULC and water quality at a catchment scale and the influence of specific LULC changes on water quality in specific region area (Li et al., 2022; McDowell, 2021), and (3) integration of several data sources including satellite images and field sampling data aimed at providing a robust approach of the linkages between LULC changes and water quality (Gani *et al.*, 2023). Several methods have been employed to explore the complex connections between LULC and surface water quality parameters. Methods that have been widely used to explore the complex connections between LULC and surface water quality parameters include lumped variables, multivariate statistical analysis, landscape pattern analysis, and scenario modelling (Gorgoglione et al., 2020; Liberoff

et al., 2019; Nguvulu et al., 2021; Ullah et al., 2018; Yao et al., 2023). However, these traditional statistical approaches have a notable limitation: they fail to consider the geographic distribution of water quality parameters, a crucial factor in comprehending water pollution dynamics. The absence of spatial information integration in analysis impedes pinpointing localised pollution sources, tracing pollutant migration patterns, and devising precise remediation strategies. Additionally, these methods share another critical limitation as they overlook the crucial aspect of Tobler's first law of geography, which states that "everything is related to everything else, but near things are more related than distant things" (Miller, 2004). This means that objects that are geographically close to each other are more likely to be similar or have some kind of spatial relationship compared to objects that are farther apart. Thus, the aforementioned methods do not explicitly account for the influence of distance and location on the relationship between LULC and water quality (Liberoff *et al.*, 2019; Lei, Wagner and Fohrer, 2022). It should be noted that the status of water quality at a given location in a river conform to the Tobler's first law of geography as Hurley & Mazumder (2013) and Liberoff et al. (2019) observed that the status of water quality at a given location in a river is influenced by the upstream watershed area, and that the influence of a specific site within that upstream area is inversely related to its distance from the point of interest. Therefore, recognising this relationship is essential when assessing the influence of LULC on water quality as it underscores the significance of considering the spatial distribution and proximity of LULC pollution sources to the point of interest (Liberoff et al., 2019).

To address this issue, spatial statistical techniques were introduced as a result of progress in remote sensing and geographical information systems (GIS) techniques (Ullah et al., 2018; Cheng et al., 2022). The spatial statistical methods such as the parametric weighted distance functions that explicitly incorporate distance and location factors, should be employed in assessing the effects of LULC on surface water quality. A parametric weighted distance function is a mathematical algorithm that measures distance by incorporating adjustable parameters to control the influence of distance on the function's output (Liberoff *et al.*, 2019). Previous research that factored in differences in landscape layout by incorporating parametric weighted distance functions, resulted in enhanced predictive capabilities in models compared to other approaches (Peterson *et al.*, 2011; Grabowski, Watson and Chang, 2016; Liberoff *et al.*, 2019).

The commonly used parametric weighted distance functions to illustrate the diminishing influence of the LULC on water quality, that is a gradual decrease in influence as distance increases include exponential and inverse distance weighted functions (King *et al.*, 2005; Van Sickle and Burch Johnson, 2008; Liberoff *et al.*, 2019). This study used the weighted inverse

distance function (WIDF) to establish the relationship between LULC and surface water quality because models that use inverse weighting function are more suitable for riparian areas, as these typically have a greater impact on stream conditions compared to locations farther away from the stream due to the inverse function's ability to create a gradual weight transition close to the stream (King et al., 2004; Peterson et al., 2011; Van Sickle & Burch Johnson, 2008). However, to the best of my knowledge, this method has not been applied in data-scarce wetland areas at the sub-catchment level like the Bangweulu Wetland.

The Bangweulu Wetland and its neighboring regions make up the lower section of the Congo River system and are acknowledged as one of the most crucial wetlands globally. Its major lakes are exploited more as a source of fish than for their tourist potential. The major lakes are known to harbor 83 species of fish and are surrounded by extensive swamps that are home to a variety of vegetation and wildlife (Ngoma *et al.*, 2017; Chundu *et al.*, 2024). The nearby areas are made up of areas where streams empty into the Bangweulu Wetland. Bangweulu Wetland in Zambia is a distinctive and crucial ecosystem, home to a diverse range of fish and wildlife (Jeffery *et al.*, 1986; Mccann, 2017; Chundu *et al.*, 2024). Nonetheless, Chundu et al. (2024) noted that there has been a swift transformation in LULC in this wetland and its surrounding areas starting from 1990 to 2020.

2.9 Research Gaps

The following gaps and limitations were identified:

- There is limited literature addressing LULC changes in the BWS.
- Understanding LULC dynamics is challenging due to the variety of classification techniques, each with its own strengths and weaknesses (Polikar, 2006; Rokach, 2010; Maxwell., 2018; Talukdar *et al.*, 2020).
- There is limited literature on rapid assessment of multiple parameters simultaneously in open water bodies, which presents significant challenges, leading to gaps and limitations in the data.
- The WQI, which integrates various water quality parameters has not been thoroughly evaluated for spatial applications (Chidiac *et al.*, 2023).
- Traditional methods for assessing the impact of LULC on surface water quality are constrained by limited spatial data coverage (Cheng *et al.*, 2022).
- There is insufficient literature on the influence of LULC on WQI, as well as limitations in the available assessment methods.

- ❖ Most commonly used approaches to determine the effects of LULC on water quality fail to consider the spatial distribution and proximity of LULC pollution sources to the points of interest (Liberoff *et al.*, 2019).

Therefore, this research used a hybrid approach of combining remote sensing and field measured data where a hybrid machine learning model was used to enhance the accuracy of modelling LULC changes in wetland areas, develop an integrated remote sensing approach for rapid holistic monitoring of water quality in open water bodies and explore the interconnected relationship between LULC and water quality status using WIDF.

2.10 Theoretical Framework

A theoretical framework is a structured system of concepts and theories that guides research by explaining the relationships among variables and providing a foundation for understanding a specific problem. It serves as a research roadmap, helping to justify and contextualise the study within existing knowledge (Hughes, 2019). This study proposes a theoretical framework based on three key concepts:

1. No Free Lunch Theorem, which states that “there is no universal machine learning algorithm that performs best on all possible problems” (Wolpert and Macready, 1997). Therefore, combining multiple ML models can enhance the accuracy of LULC classification by leveraging the strengths of different algorithms.
2. The total sum of the concentrations of water quality parameters in a water body is equal to one ($\sum Wn = 1$) (Brown et al., 1970). Therefore, evaluating the spatial variability of WQI can provide insights into the distribution of water quality across the study area thereby facilitating the comprehensive and rapid evaluation of water quality variability of open water bodies at large scale.
3. Tobler's first law of geography, which states that "everything is related to everything else, but near things are more related than distant things" (Miller, 2004). This suggests that LULC is related to the water quality status of a specific location and nearby LULC types have a greater influence than distant LULC types.

The incorporation of multiple machine learning models acknowledges the diversity of challenges inherent in environmental research, while the spatial application of WQI approach allows for a comprehensive and rapid evaluation of water quality conditions in open water bodies. Furthermore, recognising the spatial relationships articulated by Tobler's law emphasizes the importance of local LULC dynamics in influencing water quality outcomes.

CHAPTER THREE: STUDY AREA

This section provides a detailed description of the study areas, focusing on their physical and socio-economic characteristics. It also covers various aspects, including climate, hydrology, geology, soils, vegetation, and land use.

3.1 Physical characteristics of the Bangweulu wetlands

Location and Size

The eight Zambia's wetlands of international importance are situated in seven river basins: Upper Zambezi, Middle Zambezi, Lower Zambezi, Luapula, Kafue, Luangwa, and Tanganyika River basins. However, Zambia is drained by two major river systems: the Zambezi (Zambezi, Luangwa, and Kafue river basins) in the south, which covers roughly three-quarters of the country, and Luapula, Chambeshi, and Tanganyika river basins in the north, which covers roughly one-quarter of the country (Ngoma *et al.*, 2017; Ren *et al.*, 2021). This research was conducted from the Bangweulu Wetland of the Luapula River Basin. The Bangweulu (which means "where water meets the sky") wetland system (BWS) is located in Zambia's Upper Congo River Basin between 10° 45' S, 29° 30' E and 12° 40' S, 30° 40' E at an elevation of approximately 1,158 meters above sea level (Kamweneshe, Beilfuss and Morrison, 2003).

The Bangweulu swamps cover parts of Zambia's Luapula and Northern provinces, and the basin's southernmost extremity covers a small portion of the Central Province. The majority of the Bangweulu Swamps are located in the Province of Luapula. Samfya, on the shores of Lake Bangweulu, is the nearest large town. To the northeast, the area includes a portion of Isangano National Park, to the east, Bangweulu Game Management Area (GMA) number 26 and Chambeshi GMA number 27, and to the south, Kafinda GMA. Lavushi Manda National Park is fifty kilometers to the southeast of the site, and Mansa GMA number 31 is twenty kilometers to the southwest (Zambia Wildlife Authority, 2006). The wetland covers an area of about 30,000 km^2 with over 11,900 km^2 of seasonally flooded plains and permanent swamps (Kamweneshe, Beilfuss and Morrison, 2003). The wetland is fed by 17 major rivers from a catchment area of 190,000 km^2, but it is only drained by one river, the Luapula River (Kamweneshe, Beilfuss and Morrison, 2003). The study area map is shown in Figure 11.

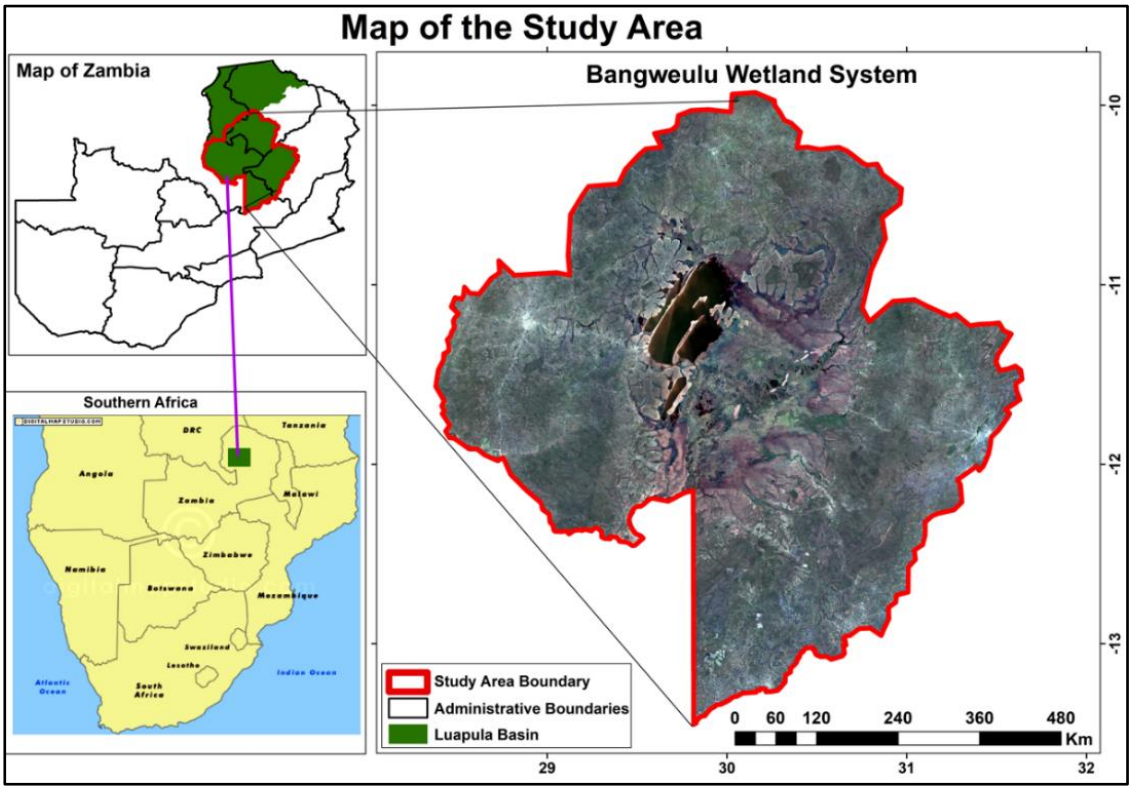


Figure 11: The location of the Bangweulu Wetland System and the surrounding areas in Zambia. The red boundary shows the exact extent of the study area, which is a sub-catchment of the Luapula Basin.

Hydrology

The seasonal flooding system, which affects every part of the local economy and environment, is the most prominent feature of the Bangweulu system. Large areas of the Bangweulu floodplains are flooded by local rainfall and runoff by January each year. Maximum water levels occur in March when the Chambeshi and other rivers draining the surrounding drainage spill across the seasonally flooded plains and permanent deep-water swamps. During the dry season, the water level gradually falls away, reaching its lowest point in November, leaving only the central basin permanently submerged (Kamweneshe, Beilfuss and Morrison, 2003). The wetland acts as natural flood control. It also aids in groundwater recharge and water quality control through plant primary productivity and sedimentation (Zambia Wildlife Authority, 2006).

BWS is one of the world's great wetland systems and comprises six interconnected lakes (Bangweulu, Kampolombo, Kangwena, Walilupe, Chali, and Chifunabuli) as shown in Figure 12. The lakes have a depth range of 3 to 10 meters and an average depth of 4 meters, with Lake Bangweulu being the largest, measuring 72 kilometers long and 38 kilometers wide (2,736 Km^2) (Hughes and Hughes, 1992; Coesel and van Geest, 2014).

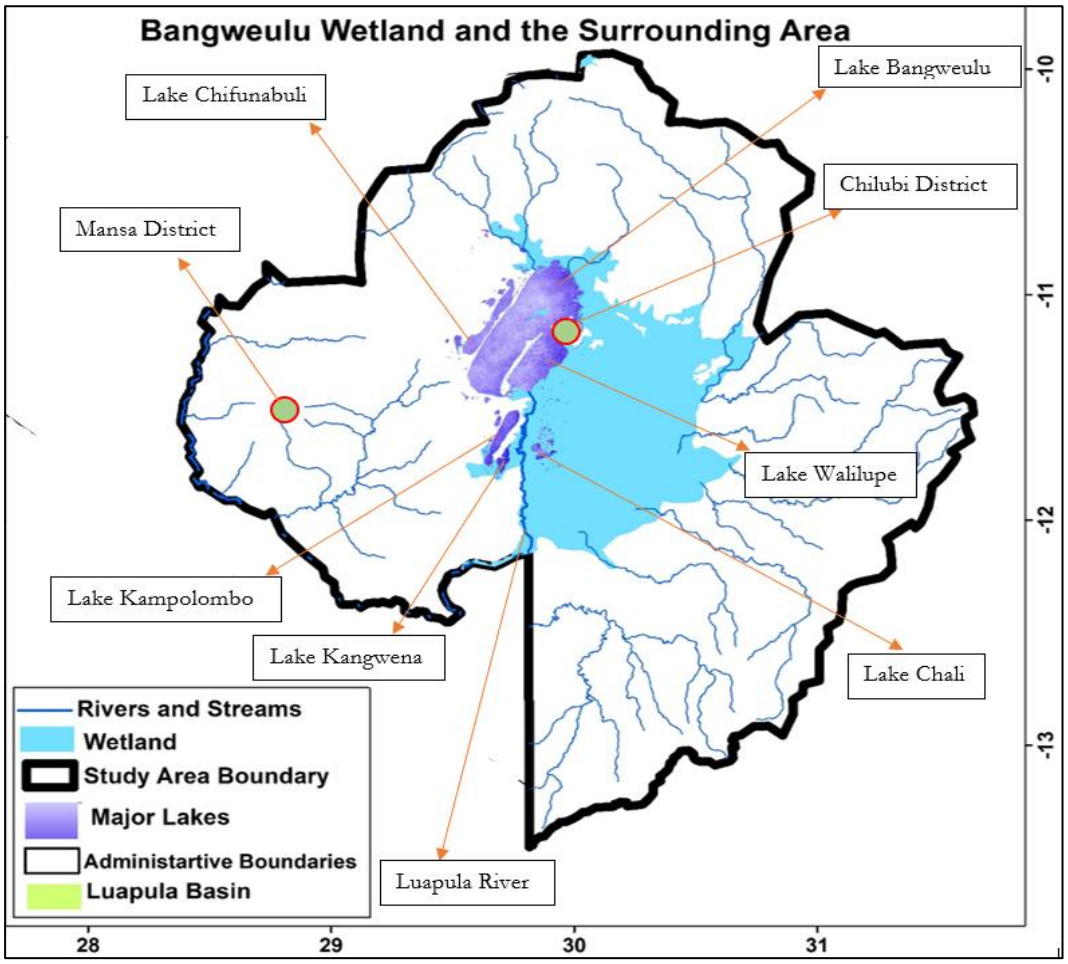


Figure 12: The Bangweulu Wetland System showing streams, rivers and the location of the six interconnected lakes.

Geology

The Bangweulu wetlands are underlain by granites, and it consists of alluvial sands and clays from river valleys, as well as sandstone, quartzite, and conglomerates which dominate the basin as a whole (Grimsdell, 1975; Zambia Wildlife Authority, 2006).

Soils

The soils of the Bangweulu wetlands are complex, with alluvial deposits predominating. In the organic topsoil, the soil is medium acidic and becomes progressively acidic with depth. The saturation base has lower calcium, magnesium, and potassium levels than most adjacent plateau soils. Large areas of peat soils similar to dambo peat are likely to be found in the deeper swamp areas. The soils in the relatively higher parts all have dark peaty topsoil (15 cm) that is mostly made up of decomposing wildlife droppings and trampled grass. The subsurface layer is 15-30 cm thick and varies in texture from loamy to clay, becoming heavier with depth. In most soils, the subsoil color is dark grayish-brown to gray, with varying degrees of mottling of yellow, brown, or red (Sen and Janssen, 1989; Zambia Wildlife Authority, 2006).

Vegetation

The Bangweulu wetland can be divided into five distinct vegetation zones: 1) Upper mainland woodland; 2) Fringing open woodlands; 3) Termitaria and fringing grasslands; 4) Seasonal floodplain; and 5) Water permanent swamp (Zambia Wildlife Authority, 2006). The upper mainland woodlands are made up of miombo trees and other related species. The trees are widely spaced, and there is a grassy understory. A termitaria zone (different habitats with termite mounds) is mostly found between open woodlands and seasonal floodplains. The termite mounds rise above the floodplain's peak flood levels, creating islands of the dense tree and shrub growth. Submerged and floating macrophytes thrive in the open waters. Grass dominates the seasonally inundated areas at the back of the landward margins of the papyrus swamps. Semi-floating swards of grasses and sedges cover the deep-water floodplain areas, including permanently inundated areas (Fanshawe, 1971; Storrs, 1995; Zambia Wildlife Authority, 2006).

Land-Use and Importance

The BWS is one of Africa's most significant ecosystems and a highly prioritized sub-catchment for protection due to its ecological and economic importance. The wetland is treeless except for private plots in the village; the dry land, particularly in the uninhabited hinterlands is densely forested (Kamweneshe, Beilfuss and Morrison, 2003; Zambia Wildlife Authority, 2006; Lehner *et al.*, 2021). The Land-use is mainly agriculture, fisheries, settlements, and wildlife. The wetlands provide not only for wildlife, such as the endemic black lechwe and hundreds of bird species but also for the 50,000 human residents (African Parks, 2021). Permanent human settlements in swamps and floodplains are mostly confined to higher ground on islands and riverbanks. Except for the scattered fishing camps that are abandoned and rebuilt each year, the rest of the area is almost devoid of human settlements. The Bangweulu fisheries are among the largest in Zambia, this has resulted in some of the highest population densities around the Lake, particularly in areas where commercial fishermen have settled. People have always lived on the outskirts of the swamp area for a long time because it has always provided a plentiful source of food. Samfya, the largest town on Lake Bangweulu, began as a fishing village in the mid-1900s (Kamweneshe, Beilfuss and Morrison, 2003; Zambia Wildlife Authority, 2006). However, the BWS is facing similar challenges of ecological degradation as a result of excessive exploitation of ecosystem services by local populations, unsustainable development and fishing practices, deforestation, and agricultural expansion (Kamweneshe, Beilfuss and Morrison, 2003; Zambia Wildlife Authority, 2006; Zambia Environment Outlook Report 4, 2017).

Climate

The Bangweulu wetland's climate is divided into three distinct seasons: a cool dry season from April to August, a hot dry season from August to October, and a warm wet season from November to April. Temperatures in the Bangweulu region range from 19^{o}C - 36^{o}C in October to 0^{o}C - 21^{o}C in July. The rainfall belt is classified as region III; average annual rainfall is around 1,100-1,500mm, with significant year-to-year variability. The annual evaporation rate is more than 2,100 mm (Kamweneshe, Beilfuss and Morrison, 2003; *Luapula Provincial Administration*, 2022).

3.2 Socio-economic Characteristics

Bangweulu Wetlands is unique in that it is made up of Game Management Areas as well as Important historical sites that can be found on its outskirts (*Bangweulu | African Parks*, 2021). The historical sites include: 1) a place where David Livingstone passed away, and his heart was interred at Chitambo. 2) The Nachikufu Caves, a site associated with the Lunda Kingdom's history. It has many wall paintings of Bushmen who are thought to have lived there before being displaced by the Lunda nation. As a result, the region is of significant historical interest and has the potential to attract both historians and scholars as well as vacationers. Each district in the Bangweulu wetlands has a different ethnic group which may have two or three ethnic groups (tribes) that speak Bemba dialects (Zambia Wildlife Authority, 2006; African Parks, 2021; Luapula Provincial Administration, 2022).

3.3 Justification for selection of study area

The Bangweulu sub-catchment was chosen for this study due to its alignment with the project's objectives, which focused on understanding the effects of Land Use/Land Cover Change (LULCC) on surface water quality. The region's unique ecological and hydrological features, including its seasonal flooding regime (Kamweneshe, Beilfuss and Morrison, 2003; Zambia Wildlife Authority, 2006; Lehner *et al.*, 2021), make it an ideal location for studying LULCC impacts. Additionally, it is a biodiversity hotspot with diverse flora and fauna, and supports vital economic activities like artisanal fisheries and tourism (Kamweneshe, Beilfuss and Morrison, 2003; Zambia Wildlife Authority, 2006; Lehner *et al.*, 2021). Despite its significance, there is a lack of prior research on LULC change and water quality in this area, making this study the first comprehensive analysis. By investigating this region, the study aims to provide valuable insights into sustainable resource management and conservation in one of Africa's most ecologically significant wetland systems.

CHAPTER FOUR: METHODOLOGY

This section provides a detailed overview of the data collection and analysis approaches used for each specific object of study. It encompasses research design, data sources, data collection methods and analysis techniques. Each methodology is tailored to address the unique requirements and characteristics of the respective study objects, ensuring a comprehensive approach to data collection and analysis.

4.1 Philosophical Orientation of this study

Philosophical orientation refers to the underlying beliefs and assumptions that guide a researcher's approach to inquiry. It shapes how researchers conceptualize their subject matter, influencing their choice of research questions, methods, and interpretations. This orientation is rooted in philosophical foundations such as ontology (nature of reality), epistemology (nature of knowledge), and axiology (role of values), which together determine what is considered valid knowledge and how it can be obtained (Kivunja & Kuyini, 2017). Therefore, the philosophical orientation of the study on the BWS embodies a realist ontology, positivist epistemology, and sustainability-oriented axiology, emphasizing the interconnectedness of humans and ecosystems while leveraging empirical methods to address pressing environmental challenges.

4.1.1 Ontological Orientation

The study adopts a realist ontology, recognizing the existence of wetlands, LULC, and water quality parameters as objective entities that interact in measurable ways. It views these entities as part of a complex socio-ecological system where human activities (e.g., land use changes) impact natural processes (e.g., water quality). This aligns with an understanding that wetlands are dynamic systems influenced by both natural and anthropogenic factors, as emphasized in ecological research. The study also implicitly reflects an interconnected ontology, where human actions and ecological systems are interdependent, resonating with relational perspectives on wetlands (Horwitz, 2022; Krueger & Alba, 2022).

4.1.2 Epistemological Orientation

The study employs a positivist epistemology by relying on empirical data, machine learning models, satellite imagery, and statistical analyses to investigate LULC changes and their impacts on water quality. This approach prioritizes observable, measurable phenomena and aims for objective knowledge through quantitative methods. The use of advanced tools like remote sensing and GIS reflects a commitment to precision and replicability in knowledge production. However, the integration of spatial distribution data and field measurements also suggests an

openness to interdisciplinary methods, bridging natural and social sciences for a comprehensive understanding of wetland dynamics (Alikhani et al., 2021; Krueger & Alba, 2022).

4.1.3 Axiological Orientation

The study demonstrates an axiological commitment to sustainability and conservation. It values wetlands for their ecological services (e.g., water quality, biodiversity) and their role in supporting human health, economic activities, and ecosystem integrity. By focusing on actionable outcomes like improved LULC policies and conservation strategies, the research reflects ethical responsibility toward preserving wetland ecosystems for future generations. This aligns with broader calls for valuing wetlands not only for their economic benefits but also for their intrinsic ecological and cultural significance (Alikhani et al., 202).

4.2 Data collection and analysis

4.2.1 Specific objective 1: To ensemble a superiorly hybrid machine learning model for enhanced accuracy of modelling LULC changes.

4.2.1.1 Field and Spatial Data

Several studies highlight the significance of having a sufficient number of training samples and the general pattern of enhanced classification accuracy as the sample size increases. However, these studies do not explicitly offer specific conclusions or optimal recommendations for sample sizes in LULC classification tasks as the appropriate sample size is determined by various factors, including the type of model used, predictor variables, LULC class definition, as well as the size and spatial characteristics of the study area (Ma *et al*., 2017; Hernandez *et al*., 2020; Mohammadpour, Viegas and Viegas, 2022).

Ramezan *et al*. (2021) investigated the impact of sample size on various machine-learning models. The results revealed a minimal decrease of 1% in overall accuracy when the training sample size ranged from 315 to 10,000 samples. Moraes *et al*. (2021) also conducted a study on the impact of sample size (50 to 6,000 samples) in LULC classification using Random Forest and found that 2000 samples achieved the highest accuracy (73.7%), while 6000 samples yielded the lowest accuracy (71.5%). Despite a drastic reduction in training sample units, there was only a 2% accuracy variation, indicating consistent classification accuracy performance. Furthermore, different studies have used different numbers of training samples. For example, Kulkarni and Lowe (2016) trained the Random Forest model using the dataset of 600 samples and achieved an overall accuracy of 96%. Laban et al. (2019) utilized 708 training and validation samples for various machine learning LULC classifications, resulting

in an overall accuracy rate of 92.7%, 92%, 92.1%, and 94.4% for the RF, K-NN, ANN, and SVM algorithms, respectively.

In view of the above, this research considered the use of cost-effective methods by using minimum training samples that could still yield desirable performance for modelling LULC changes. Therefore, I employed purposive sampling, wherein the researcher selected elements to be included in the sample based on their relevance to the study topic and the prevailing conditions in the field as described by Etikan (2016). We used a combination of field surveys conducted between October to December 2022 and March to April 2023, and a visual inspection of very-high-resolution images available on Google Earth Pro to collect training and validation data for the 2020 image. This approach allowed us to create a total of 800 polygons for training and validation purposes.

During the field investigations, 315 geometry (Appendix II) points were collected from accessible areas representing various LULC categories, namely built-up areas (including roads, settlements, and barren land), grasslands (all grass types and open shrubs), forests (encompassing thick shrubs and wooded regions), croplands, and water bodies. The geometry points were superimposed on the 2020 Landsat image and 315 polygons were created based on the location of geometry points.

Additionally, the 315 training and validation polygons were supplemented with 485 other polygons. This supplemental dataset of polygons was created through visual inspection of high-resolution images obtained from Google Earth Pro, which represented various LULC classes and served as an additional source of information for training and validating the models. To obtain the coordinates and attribute information of LULC, I primarily relied on a handheld global positioning system (Garmin GPS). We also relied on a visual inspection of the Google Earth pro time series, the Landsat images, and the researcher's expertise to create the training and validating dataset of polygons for 1990, 2000, and 2010 images.

4.2.1.2 Data Partitioning

The dataset containing 800 polygons was partitioned into five segments based on the following percentages: 60%, 10%, 10%, 10%, and 10%, resulting in 480, 80, 80, 80, and 80 polygons, respectively. The 60% portion and one of the 10% portions were allocated for model development and validation, respectively. The remaining three 10% portions were set aside specifically for conducting triple post-classification accuracy assessment, as illustrated in Figure 13. A triple cross-validation was used to enhance the reliability of the model's LULC classification and the averaged results obtained through cross-validation were reported as the

main performance of the model. Reporting averaged metrics enhances transparency and facilitates fair comparisons among different models. These results provide a more robust and representative evaluation, accounting for the variability of the data and offering a reliable estimate of the model's performance on unseen samples (Basheer et al., 2022; Hosseiny, Abdi and Jamali, 2022). Figure 13 shows the distribution of training and validation polygons in the study area.

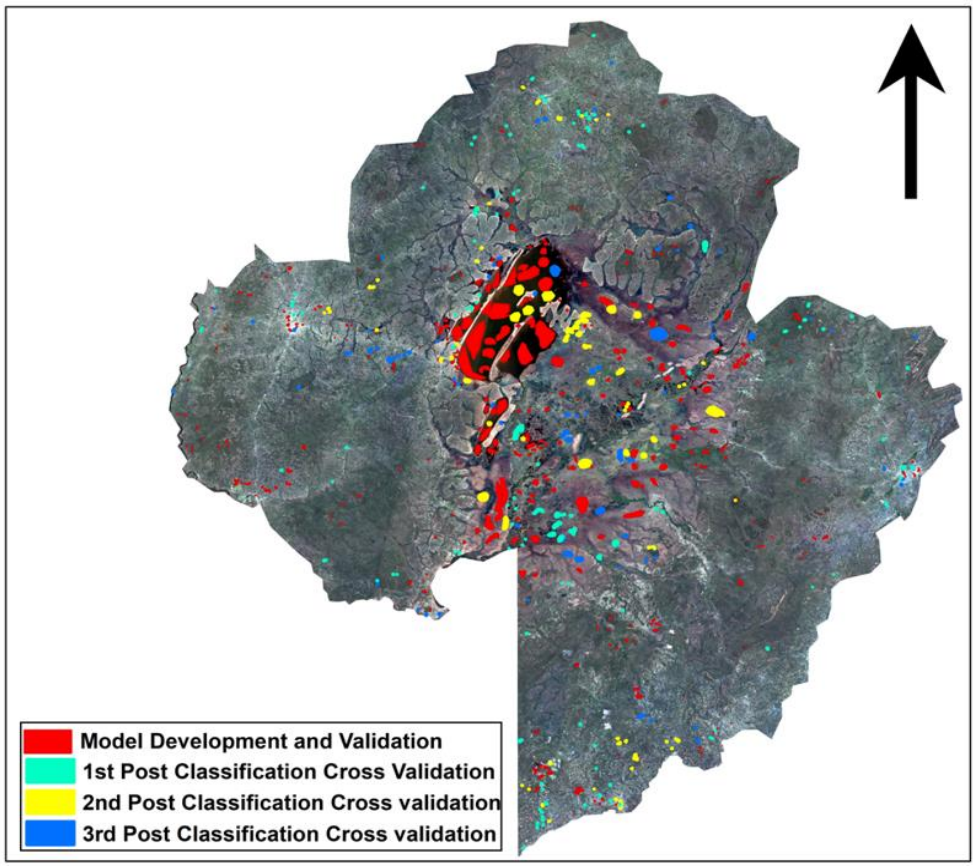


Figure 13: Distribution of the model development, validation, and post-classification cross-validation polygons for 2020 Landsat image based on field data and Google Earth Pro. For the red polygons (560 polygons), the algorithms picked at random 480 and 80 polygons for model development and validation, respectively.

4.2.1.3 Spatial data

Annual average conditions of pre-processed images of Landsat 5 (1990, 2000, and 2010 images) and Landsat 8 (2020 image) with four bands (near-infrared, red, green, and blue) of the Bangweulu Wetland and the surrounding areas were downloaded from Climate Engine (https://app.climateengine.com/climateEngine). The annual average conditions of pre-processed images are single average images generated by the Climate Engine by combining several input images from January to December (Huntington *et al.*, 2017). The downloaded images were classified using various Machine Learning Algorithms (ANN, NB, SVM, DT, RF, and KNN algorithms) through the Orfeo Toolbox in QGIS 3.28.5.

4.2.1.4 Model Development Set parameters

The parameters for various algorithms used in the model development were established by taking into consideration the guidelines of various scholars as well as fine-tuning their specific parameters through trial and error thereby ensuring that each model achieved the best possible performance. For RF, the following number of trees were used for fine turning 10, 50, 100, and 150. However, different numbers of trees did not provide sufficient influence on the classification results. This observation was expected as it has already been reported by other scholars (Guan *et al.*, 2013; Belgiu and Drăgu, 2016). Therefore, the default setting value (100) of the Orfeo Toolbox in QGIS was used. Similarly, the default value (32) of ***K*** was also used for KNN. For SVM and DT, the parameters used were based on the findings of various scholars (Gholami and Fakhari, 2017; Abidi *et al.*, 2020; Ramezan *et al.*, 2021). The full list of the parameters used are presented in Table 5.

4.2.1.5 Model Development and Accuracy Assessment

Various algorithms, namely ANN, NB, SVM, DT, RF, and KNN were used for model development. The models were created using a total of 480 reference polygons as training data samples, and an additional 80 reference polygons were reserved for evaluating the accuracy of the models. Among the developed models, those that demonstrated exceptional performance, surpassing the others with a Kappa Index (KI) exceeding 0.60 were selected for further analysis based on Table 5.

Table 5: The specific input parameter values for different models

	Model	Parameters			
1	SVM	Kernel type RBF	Cost 4	Gamma 0.033	SVM type CSVC
2	DT	Maximum depth of Trees 10	Minimum number of samples in each node 10	Random seed 0	
3	RF	Number of trees in forest 100	Random seed 0	-	-
4	KNN	Number of Neighbours 32	-	-	
5	NB	-	-	-	-

*Assumption: The parameters used in Table 5 were assumed to be optimal.
RBF: Radial Basis Function kernel, CSVC: Complex-valued Support Vector Classifier.

4.2.1.6 Image Classification and Triple Cross Validation

The selected models were employed to classify images from the years 1990, 2000, 2010, and 2020. The resulting classified maps underwent triple cross-validation using three distinct sets of 80 reference polygons. Cross-validation is a widely used procedure for validation of the LULC classification models. One subset of the dataset is used for model training, and another subset is used for validation. This is done for multiple (k-fold) times, and the results are eventually averaged to get a more robust assessment of the model's performance (Hayes, Miller and Murphy, 2014). In this research post classification model validation was done using three different datasets. Therefore, Triple (3-fold) cross-validation was the name given to the three-step cross-validation process, and the averaged Kappa Index (KI) was used to select a more accurate model's performance estimate. This accuracy evaluation process was referred to as post-classification triple validation. The average value of the KI obtained from this assessment was used to identify the models with the highest performance. Models that had a triple cross-validation KI average below 0.60 were excluded, while those with a KI exceeding 0.6 were selected for further analysis, this was based on Table 6.

4.2.1.7 Fusion of Classified Maps and Triple Cross Validation

The classified maps from the best-selected models were fused using majority voting mechanism in QGIS - Orfeo Toolbox fusion of classifications, to create a hybrid map representing the LULC classes for the years 1990, 2000, 2010, and 2020. The resulting hybrid maps also underwent triple cross-validation using the same three sets of 80 reference polygons. This was

done to compare the accuracy of the hybrid classified maps with the individual models classified maps.

4.2.1.8 Area changes between 1990 and 2020

The change map was realised by finding the difference between the 2020 and 1990 hybrid maps. In addition, the annual rates of LULC change (ARC) as illustrated in Equation 1, were calculated using the following formula:

ARC = ((LULC_end - LULC_start) / LULC_start) * 100 / (number of years)… Equation 1

Where:

- LULC_start is the value of LULC at the start of the period.
- LULC_end is the value of LULC at the end of the period.
- Number of years: years between the start and end of the period.

4.2.1.9 Metrics for Accuracy Assessment

The metrics employed to evaluate the accuracy of the models' performances were the Kappa Index (KI) (Equation 2), Overall Accuracy (OA) (Equation 3), and F-score (Equation 4) of the LULC classes.

Kappa index

The Kappa index (KI) was calculated as shown in Equation 2 (Zhang *et al.*, 2021a; Alshari and Gawali, 2022)

$$KI = \frac{(Po - Pe)}{(1 - Pe)} \quad \text{……. Equation 2}$$

Where:

- P_o is the proportion of agreement between LULC classes and reference datasets.
- Pe is the sum of the product of the row and column marginal proportions for each category, divided by the total number of observations squared.

KI value ranges from -1 to 1,Where:

- KI = 1 indicates perfect agreement.
- KI = 0 indicates agreement that is no better than chance.
- KI < 0 indicates agreement worse than chance.

Overall Accuracy

The OA as illustrated in equation 3, provides a measure of how accurately the model assigns the correct class to each pixel or sample. The equation sums up the correctly classified samples (both positive and negative, i.e., the sum of the values on the major diagonal) and divides it by the total number of samples in the dataset (Congalton, 2001).

$$OA = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad \text{..... Equation 3}$$

Where:

- TP (True Positive) represents the number of correctly classified pixels or samples belonging to the positive class.
- TN (True Negative) represents the number of correctly classified pixels or samples belonging to the negative class.
- FP (False Positive) represents the number of incorrectly classified pixels or samples that were wrongly assigned to the positive class.
- FN (False Negative) represents the number of incorrectly classified pixels or samples that were wrongly assigned to the negative class.
- OA ranges from 0 to 1, where 1 indicates perfect classification accuracy and 0 indicates no accuracy at all.

F-Score

The F-score (Equation 4), also referred to as the F1 score or F-measure serves as a performance metric for evaluating Machine Learning models. It combines precision and recall into a single metric by calculating their harmonic mean. This harmonic mean places more emphasis on lower values, making it a suitable metric when precision and recall need to be balanced. Precision measures the accuracy of positive predictions, while recall quantifies the model's ability to identify all positive samples. The F-score, ranges from 0 to 1, with higher values indicating better accuracy (Dalianis, 2018).

$$\text{F-score} = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad \text{..... Equation 4}$$

However, the KI in this research was the primary metric used in the elimination of underperforming models. Any model with KI less than 0.60 was eliminated. This was based on (Landis and Koch, 1977; McHugh, 2012) as depicted in Table 6.

Table 6: KI interpretation

Value of Kappa Index	**Level of Agreement**
-1 - 0.20	None
0.21 - 0.39	Minimal
o.40 - 0.59	Weak
0.60 - 0.79	Moderate
0.80 - 0.90	Strong
Above 0.90	Almost Perfect

Adapted from Landis and Koch (1977) and McHugh (2012).

Figure 14 serves as a comprehensive visual representation of all the different materials and methods we used in enhancing the accuracy of modelling LULC changes in wetland areas and how they were combined at different stages.

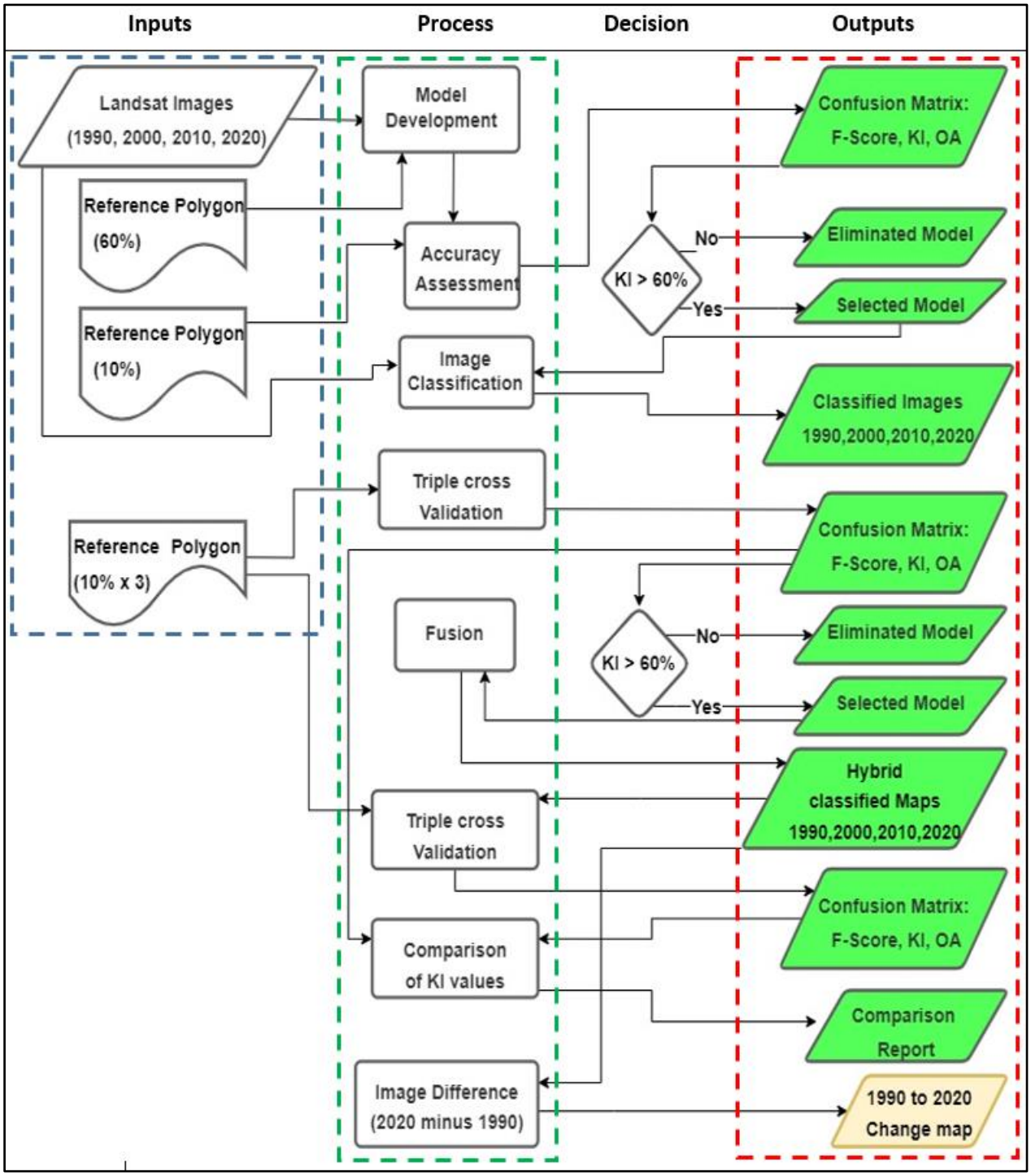


Figure 14:A summary of the research materials and method on the enhancement of the accuracy of modelling LULC changes.

4.2.2 Specific objective 2: To rapidly evaluate the variability of water quality.

4.2.2.1 Water sampling

The sampling design for the water quality assessment took into account two major criteria: Firstly, to account for potential sources of pollution, sampling points were chosen near various land uses, including built-up and agricultural areas, fishing camps, crossing spots, and trading areas. Additionally, an area with no anthropogenic activities was also selected to cover other variations in water quality. Secondly, to minimize the inclusion of mixed pixels (adjacency effect) at the edges of the lakes or where the lakes are narrow, sampling stations were selected based on the availability of an open water body with a width of more than 10 meters. Specifically, we considered at least four times the size of a sentinel pixel (10m), i.e. a 40m× 40m area, which is equivalent to 16 (4×4) pixels. This ensured that the pixel did not contain any significant signal from spectral mixing with surrounding land or terrestrial vegetation (Hestir *et al.*, 2015). Furthermore, the adjacency effect is more pronounced in dark waters compared to bright waters like that of the Bangweulu wetland lakes (Paulino *et al.*, 2022). To consider all these aspects, a non-probability sampling method was necessary. Therefore, we used a purposive sampling method to select the water sampling points. Purposive sampling is a type of non-probability approach where the researcher selects elements to include in the sample based on what is needed in the topic under study and the prevailing conditions in the field (Etikan, 2016). Water samples were collected in the dry season between October and November 2022 as well as November 2023. The coordinates of each water sampling point were collected using a handheld Global Positioning System (Garmin GPS). The choice of the period for data collection was to ensure that the data was collected when the wetland subsided to its lowest levels. This was based on both field and historical data (Figure 15).

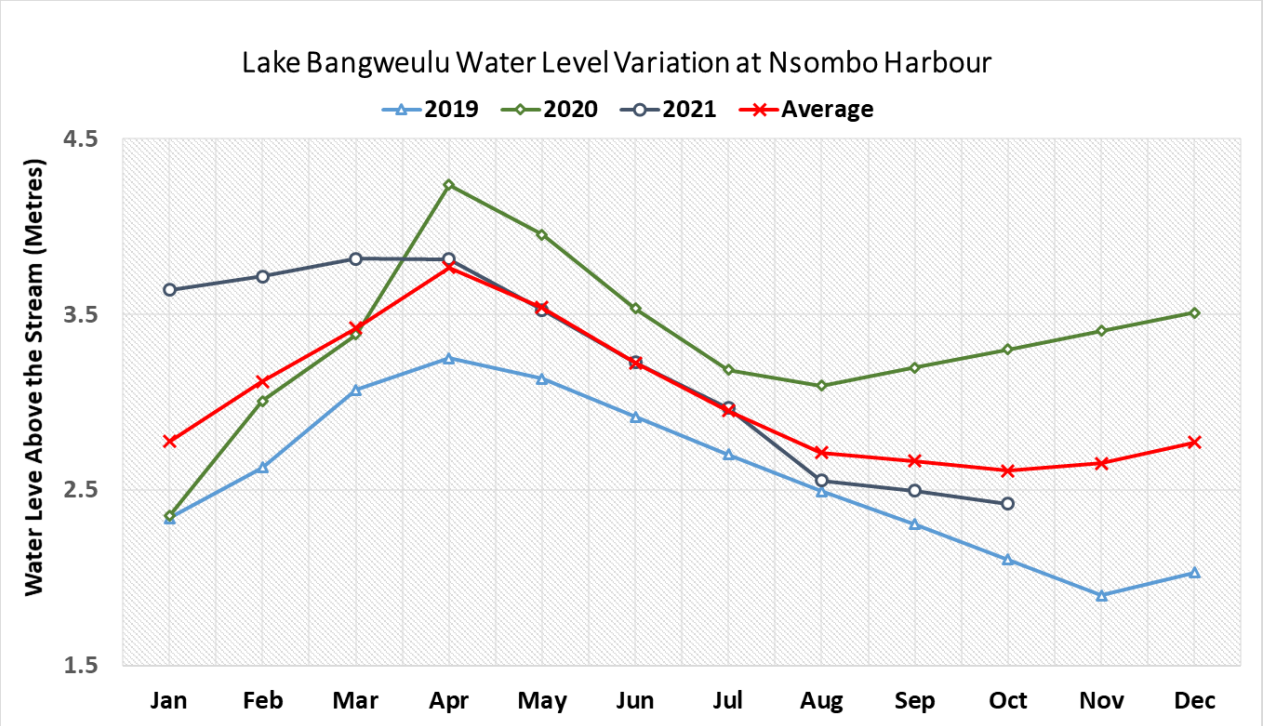


Figure 15: Shows the distribution of the water sampling points in the study area. Source: Water Resources Management Authority (WARMA), Zambia.

The on-site water quality parameters that were measured using the ProDSS Multi-Parameter water quality meter and the lab-measured water quality parameters are indicated in Table 7 and Appendix III. The laboratory analysis of the water quality was done at the University of Zambia Environmental and Geo-chemical laboratories. At each sampling point, 500 and 100 milliliters (ml) of water were collected for laboratory analysis. Subsequently, the 500 mm water samples were stored at 4 degrees Celsius to minimize biological activity before their transportation for laboratory analysis. Whereas the 100 ml water samples designated for heavy metal water quality analysis had 1 ml of nitric acid added to stabilise the heavy metal concentrations over time before being transported for laboratory analysis. The spatial distribution of the sampling points is shown in Figure 16.

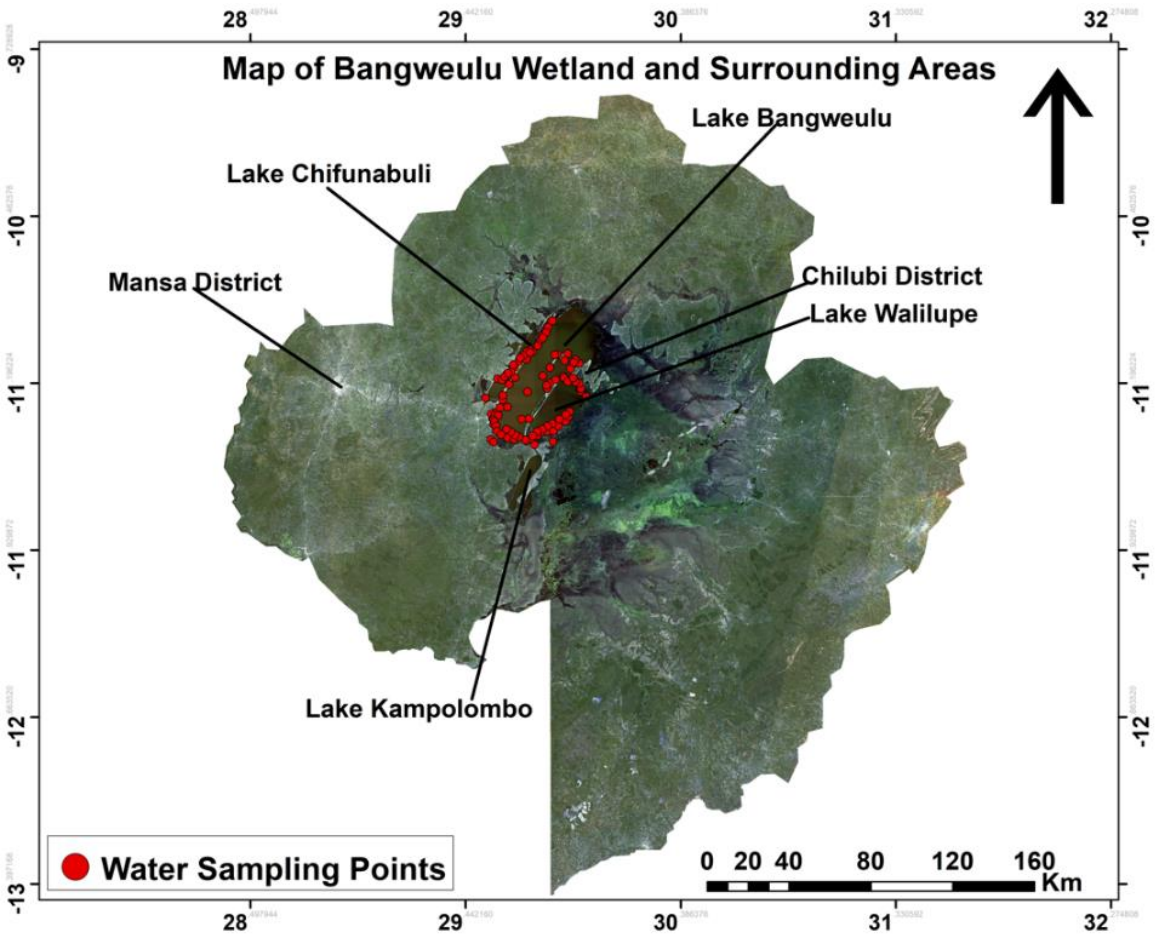


Figure 16: Map showing the major lakes of the Bangweulu Wetland and the spatial distribution of the sampling points. Water samples were collected in the dry season between October and November 2022 as well as November 2023.

The values for the on-site and lab measurements of different water quality parameters for each sampling point and the summary of the water quality parameter measurements as well as the Zambia Bureau of Standards (ZABS) ambient water quality (These are the optimal water use thresholds in the Luapula Catchment/Bangweulu Wetland lakes) and International (World Health Organisation) standards for drinking water are displayed in table 7. The ambient water quality standards were used for subsequent analysis in this research because they have the highest local standards of water purity needed for the Bangweulu Wetland.

Table 7: The table shows the summary of on-site water quality measurements, laboratory results, Zambia Bureau of Standards (ZABS) ambient standards and International Standards for drinking water.

Parameter	Lowest Value	Highest Value	Mean Value	Ambient Water Quality	International Drinking Water Standards	Sampling Points	Measurement mode
EC (µS/cm)	14	40	25.99	100	400	64	On-site
TDS (mg/l)	9	25	16.71	45	500 -1000	74	On-site
Turbidity (TNU)	0.47	19.3	2.24	10	5	74	Laboratory
Sodium (mg/l)	1.09	3.27	1.99	5	30 - 60	58	Laboratory
Chloride (mg/l)	5	15	9.15	30	250	58	Laboratory
Potassium (mg/l)	3.3	9.9	6	NIL	12	58	Laboratory
Faecal (cfu/100ml)	0	>100	28.28	50	0	93	Laboratory
Calcium hardness (mg/l)	3	11	7.61	30	10 - 500	64	Laboratory
Total hardness (mg/l)	8	18	13.44	250	100 – 300	64	Laboratory
Iron (mg/l)	0.002	0.623	0.2	0.7	0.3	49	Laboratory
Lead (mg/l)	<0.002	<0.002	N/A	2	0.01	49	Laboratory
Manganese (mg/l)	<0.002	<0.002	N/A	0.001	0.4	49	Laboratory
Nitrates (mg/l)	<0.01	<0.01	N/A	6	Oct-50	34	Laboratory
Phosphate (mg/l)	<0.01	<0.01	N/A	0.04	0.1	34	Laboratory
Sulphates (mg/l)	<0.01	<0.01	N/A	20	250	34	Laboratory
pH	6.68	8.4	7.3	5.5 – 8.0	6.5 - 9.5	38	On-site
Temp. oC	24	27	25.63	23.5	25	64	On-site

Source of water quality standards: (Frisbie et al., 2012; Hong et al., 2023; Meride & Ayenew, 2016; ZABS office, Zambia).

4.2.2.2 Spatial data

The pre-processed mean condition of sentinel 2 images from 1st to 30th November 2022 were downloaded from the Climate Engine (https://app.climateengine.com/climateEngine). The Climate Engine creates single average image representing the monthly average conditions by integrating multiple pre-processed input images for a specified period (Huntington *et al.*, 2017; Chundu *et al.*, 2024). We used average November 2022 imagery for our analysis because the in-situ water quality data was gathered at various dates within the fourth week of October to November 2022 and November 2023. This was challenging to pinpoint satellite imagery of these specific dates, and therefore, the average imagery was used based on the assumption that lakes typically exhibit stability during the summer months, with the notable exceptions of algal growth and pH level fluctuations (Lihepanyama, Ndakidemi and Treydte, 2022) Zeng et al., 2022). Additionally, we assumed that the water quality parameter concentration levels did not change significantly between 2022 and 2023 (Horvat, Horvat and Pastor, 2021; Zainurin *et al.*, 2022). The individual Sentinel 2 bands used in this study include Near Infrared (NIR), Red, Green, and Shortwave Infrared 1 (SWIR 1) (refer to Table 4 for specific details). The downloaded bands were used to calculate MNDWI, NDSI, and NDTI.

a. Modified Normalised Difference Water Index (MNDWI)

The MNDWI as expressed in equation 5 was used to extract the water bodies from the study area based on Xu (2006). The extracted water bodies' extent was converted to shapefile for further analysis.

MNDWI = (Green – SWIR 1) / (Green + SWIR 1) …… Equation 5

Where *Green* and *SWIR1* are the reflectance values of the *Green* and *SWIR1* bands, respectively. The **MNDWI index has a range of -1 to 1**, with near 1 indicating open water, close to 0 indicating land surfaces, and negative values indicating dense vegetation or built-up areas.

b. Normalized Difference Salinity Index (NDSI)

The NDSI was calculated as expressed in equation 6, and it was based on (Khan *et al.*, 2005; Al-Jabri *et al.*, 2023)**.**

NDSI = (Red − NIR) / (Red + NIR) …… Equation 6

Where *Red* and *NIR* are the reflectance values of the *Red* and *NIR* bands, respectively.

NDSI values range from **-1 to +1.** The higher positive NDSI values generally indicate higher salinity levels, whereas lower or negative values indicate lesser salinity or freshwater.

The coordinates of the field sampling points were overlaid on the NDSI map. For each sampling point, a shapefile with a 10m buffer was created. This buffer represented an area surrounding the sampling point. The next step involved the use of the shapefile with the 10m buffer to extract the NDSI zonal mean statistics for each sampling point. This meant that the average NDSI value within the 10m buffer around each sampling point was calculated. These NDSI mean zonal statistics values were then used to establish a relationship between on-site/lab-measured sodium, potassium, chloride, and EC levels in water and the NDSI. The goal was to determine if there was a correlation between these variables. Using regression analysis, an equation was derived to describe the relationship between sodium, potassium, chloride, EC levels in water, and the NDSI zonal mean values. This equation was then used to predict the sodium, chloride, and EC levels based on the NDSI values. Finally, the NDSI map was transformed into individual maps for sodium, chloride, and EC by applying the regression equation to each pixel in the NDSI map. This transformation allowed for a visual representation of the water quality parameters levels across the area covered by the NDSI map.

c. Total Dissolved Solids (TDS)

The TDS maps were created from its relationship with EC (refer to equation 7) which states that the TDS in milligrams per Litre (mg/L) is equal to a constant (k) multiplied by the EC in micro Siemens per centimetre (µS/cm) (Taylor, Elliott and Navitsky, 2018)

TDS (mg /L) = k x EC (µS /cm) ……. Equation 7

Where ***k*** is a proportionality constant. The value of ***k*** is determined by the type of solution being measured as well as its temperature. The ***k*** value is affected by the ionic composition of water and the concentration of dissolved organic and inorganic substances. The ***k*** value for most natural water ranges from 0.55 to 0.85, with an average of 0.7 usually used (Atekwana *et al.*, 2004; Taylor, Elliott and Navitsky, 2018; Walton, 1989). Therefore, setting the ***k*** value at 0.7, the linear equation (Equation 7) was then used to transform the EC map into TDS map.

d. Normalised Difference Turbidity Index (NDTI)

NDTI was calculated by subtracting the green spectrum value from the red spectrum value and dividing the results by the sum of both values (Equation 8). This computation is based on the assumption that water with great clarity reflects more in the green wavelength than in the red spectrum. The reflectance in the red spectrum increases as the water becomes more turbid (Lizcano-Sandoval *et al.*, 2022).

NDTI = (Red − Green) / (Red + Green) ..... Equation 8

Where ***Red*** and ***Green*** are the reflectance values of the ***red*** and ***green*** bands, respectively.

NDTI ranges between -1 and 1, an increase in turbidity or decreased water clarity is indicated by a positive NDTI value, whereas a decrease in turbidity or increased water clarity is indicated by a negative value. To create the turbidity map, the procedure described in subsection 3.2.2.2b above was followed. But in this case, the NDTI zonal mean values were extracted from each sampling point using the shapefile with the 10m buffer. The NDTI zonal mean statistics were then used to establish a relationship between lab-measured turbidity and the NDTI zonal mean values at the sampling points using regression analysis, an equation was derived. The equation was then used to predict the turbidity levels based on the NDTI zonal mean values. Finally, the NDTI map was transformed into a turbidity map by applying the regression equation to each pixel in the NDTI map.

4.2.2.3 The Water Quality Index (WQI)

The WQI for the on-site/lab-measured and remotely sensed water quality parameters were calculated based on the weighted arithmetic mean equation by (Brown *et al.*, 1970) (refer to Equation 9).

WQI = ∑QnWn / ∑Wn ..... Equation 9

Where:

Qn is the water quality parameter's quality rating the nth water quality parameter, and ***Wn*** is its unit weight of the ***nth*** water quality and the ***∑Wn = 1.***

The ***Qn*** was determined as follows:

Qn = (Vn-Vi / Vs-Vi) *100

Where:

Vn indicates the observed value of the parameter, ***Vi*** represents the ideal value of that parameter. ***Vi*** ***=*** *0*, except for *pH* ***(Vi = 7)*** and *DO* ***(Vi = 14.6 mg/l)***

Vs is the standard permissible value for the water quality parameter

Wn = K / Vs

The constant of proportionality ***(K)*** was computed using the following equation:

K = [1 / (∑1/Vs)]

The WQI was classified as shown in Table 8

Table 8: Water Quality Index (WQI) range, state, and potential uses of water

WQI	Water quality status	Water quality grade	Possible uses
0 - 25	Excellent	A	Drinking, irrigation and industrial purpose.
26 - 50	Good	B	Drinking, irrigation and industrial purpose.
51 - 75	Poor	C	Irrigation and industrial purpose.
76 - 100	Very poor	D	For irrigation purpose.
Above100	Unsuitable	E	Proper treatment required for any kind of usage.

Adapted from: (Brown et al., 1970; Islam et al., 2021)

To ensure adaptability to the available water quality parameters data set, this study opted for an open system approach in selecting water quality parameters of interest because the number of parameters considered in this study depended on the available water quality data obtained from the on-site/lab and through remote sensing methods. To calculate the on-site/lab-measured WQI, water quality parameters displayed in Table 7 above were used. Whereas, to calculate remotely sensed WQI, zonal mean values from a 10m buffer of each water quality parameter map were used. Subsequently, the remotely sensed WQI values from different locations were utilised to establish the relationship between the WQI and the water quality parameters with the resulting regression equations being used to transform the water quality parameter maps into WQI maps.

4.2.2.4 Reclassification and Integration of WQI Maps

The classification of the WQI maps into different classes was done based on Table 8 guidelines. To create a single integrated WQI map, a weighted overlay approach with equal weights of different WQI maps was used using ArcGIS. The summary of the whole research methodology in developing the integrated remote sensing approach for water quality monitoring is illustrated in Figure 17.

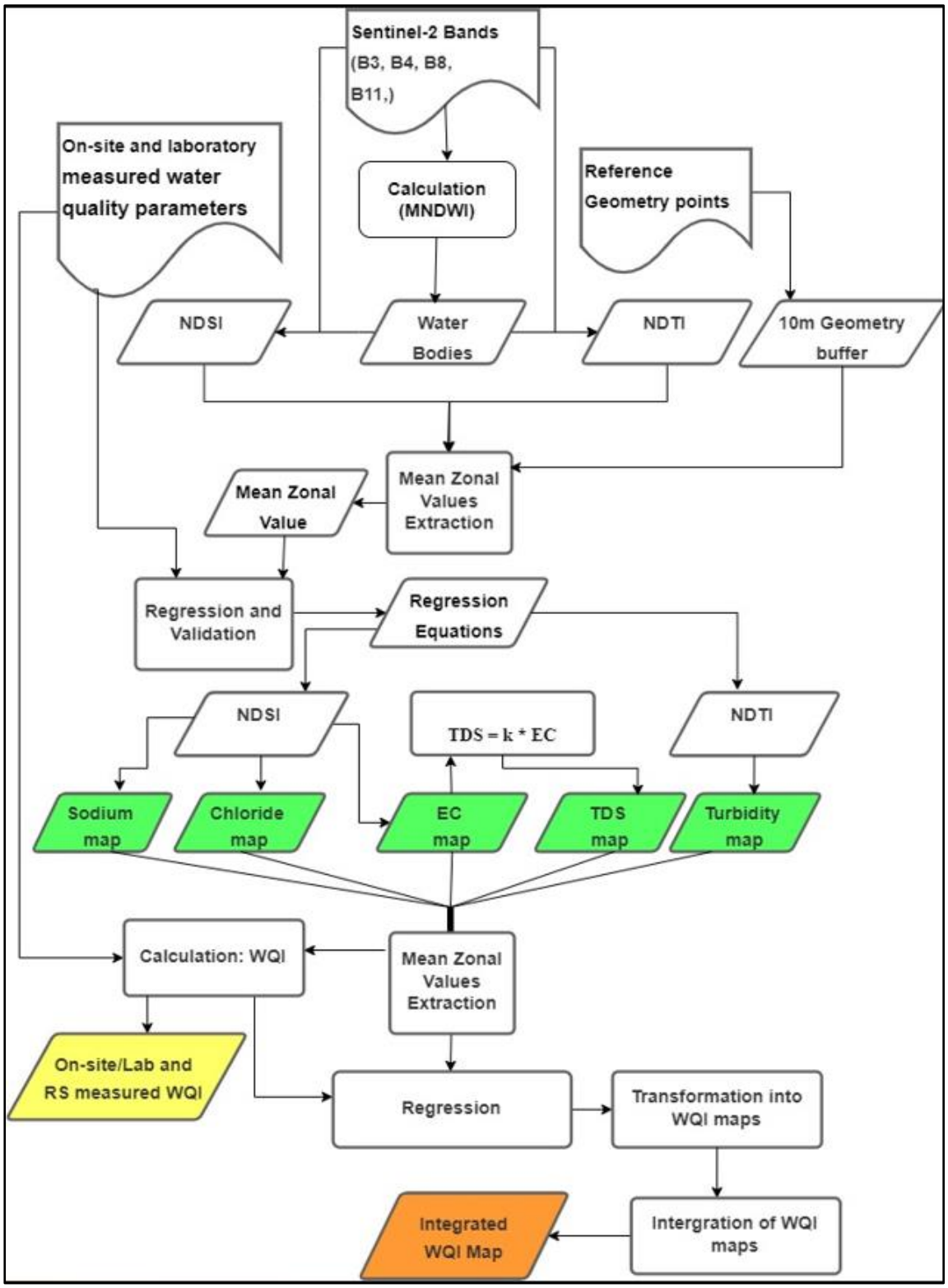


Figure 17: Summary of the research methodology on the development of integrated remote sensing approach for water quality monitoring. The green boxes are the individual water quality maps whereas the yellow and orange boxes are WQI values and integrated water quality map, respectively. For $\sum Wn = 1$, refer to Equation 9.

4.2.3 Specific objective 3: To investigate the influence of LULC on water quality.

4.2.3.1 Field and Laboratory Water Sampling and Measurements

The selection of water quality sampling stations was based on their closeness to various LULC classes. Hence, a purposive sampling approach was also utilised to choose the water sampling locations. Water samples were collected and analysed as explained in subsection 3.2.2.1. Figure 18 displays the spatial distribution of the 34 sampling points used in the assessment of the linkages between LULC and water quality and Appendix IV indicates their geometry points.

4.2.3.2 Spatial Data

This study made use of the 2020 LULC classified maps (Figure 18A) from Chundu *et al.*(2024) and the 30-meter resolution of Digital Elevation Model (DEM) (Figure 18B) obtained from the Shuttle Radar Topography Mission (https://dwtkns.com/srtm30m/).

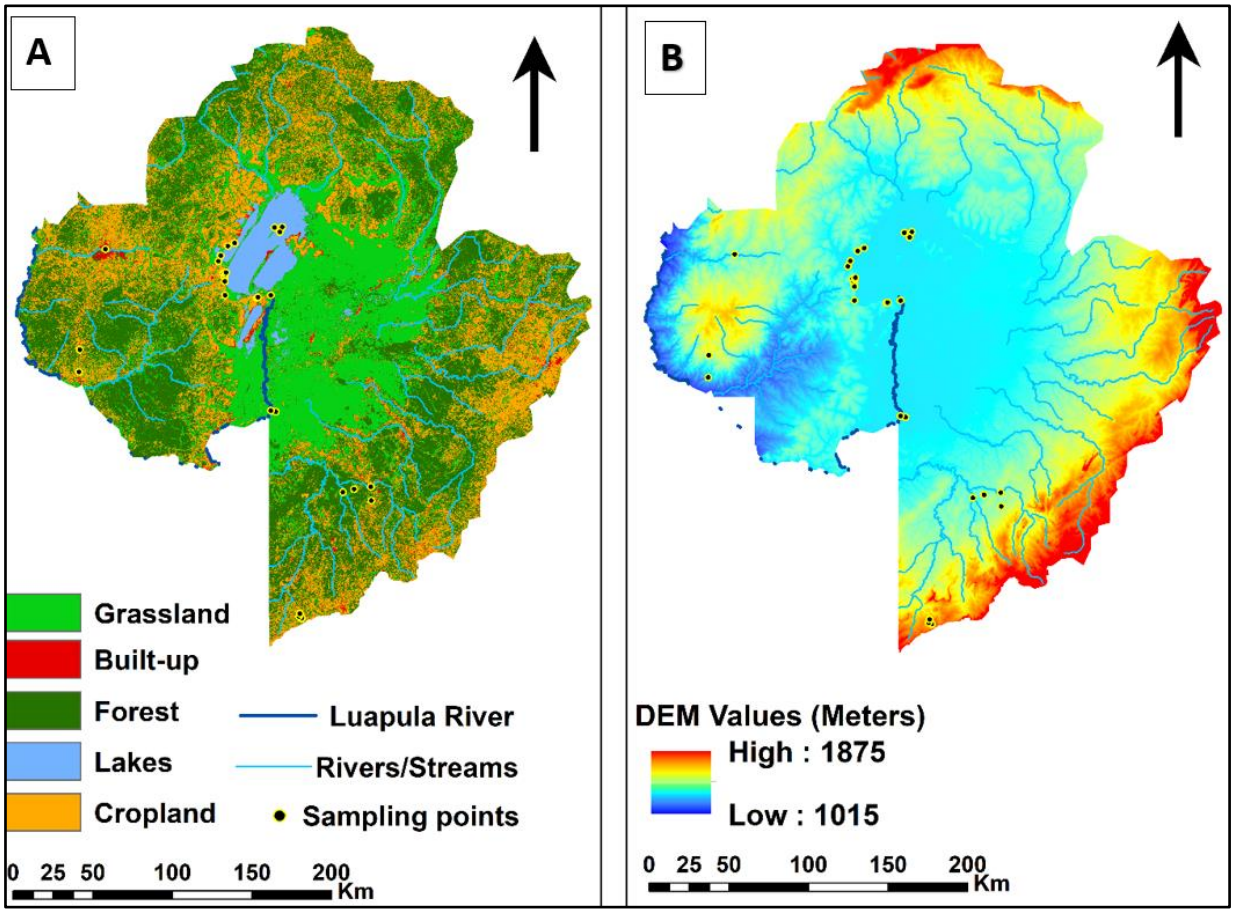


Figure 18: Spatial Data used in this study: (A) 2020 LULC classified maps and (B) 30-meter resolution of Digital Elevation Model.

4.2.3.3 Water Quality Index (WQI)

The WQI for each sampling point was calculated based on the weighted arithmetic mean equation by Brown *et al.*, (1970) as explained in subsection 3.2.2.3, equation 9 above.

4.2.3.4 Watershed Area Delineation

The watershed area was delineated using QGIS, which involved several steps. First, the 30m DEM (Figure 19A) was reprojected to WGS 84 UTM zone 35s to ensure accurate spatial analysis. Next, the Fill Sinks tool, also known as the Wang and Liu method, was used to fill any depressions or sinks (Figure 19B) in the reprojected DEM, ensuring a continuous flow of water across the terrain. After that, a Strahler Order (Figure 19C) was generated to outline the stream network, which is an essential indicator of the desired catchment and the various water networks. Then, the water sampling points were superimposed on the Strahler Order as outlet points.

The Upslope Area tool was used to generate the watershed based on the input raster layer (filled DEM) and the target x,y coordinates (sampling points). The tool calculates the upslope area (Figure 19D) for each cell in the filled DEM, which is the area that contributes water to that cell. The output is a raster layer that shows the upslope watershed area that contributes water to a specific sampling point, representing the outlet point of the watershed. To save the delineated watershed area, the raster watershed was converted to a vector shapefile format for further analysis.

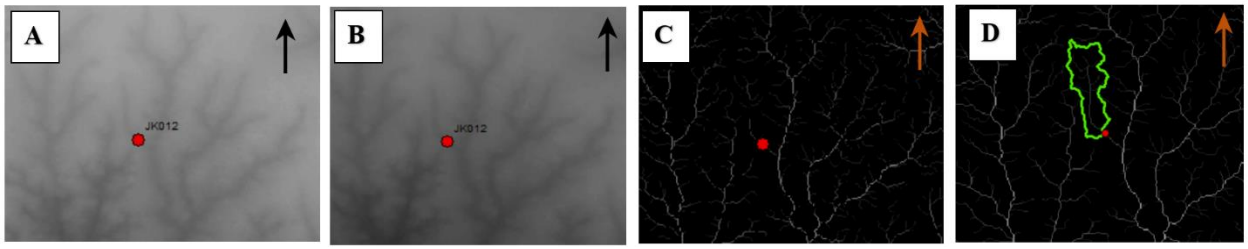


Figure 19: Watershed Area Delineation steps: (A) 30m DEM, (B) Sink filled 30m DEM, (C) Strahler Order, and (D) Watershed upslope area.

4.2.3.5 Euclidean Distance

In order to determine the Euclidean distance between two points, specific steps were followed in QGIS. This involved utilising the delineated watershed area to extract the 2020 LULC (Figure 20A) influencing the sampling point. The LULC found within the designated watershed area was converted into point features (Figure 20B), with each point representing a distinct LULC type. The Euclidean distance raster (Figure 20C) was generated by using the source point (sampling point) and the LULC

points (processing extent). Next, the LULC points layer was then overlaid (Figure 20D) on the Euclidean distance raster, and the corresponding Euclidean distance of each LULC point was extracted. This process allowed for the calculation of the straight-line distance between the sampling point and the different LULC classes within the delineated watershed area.

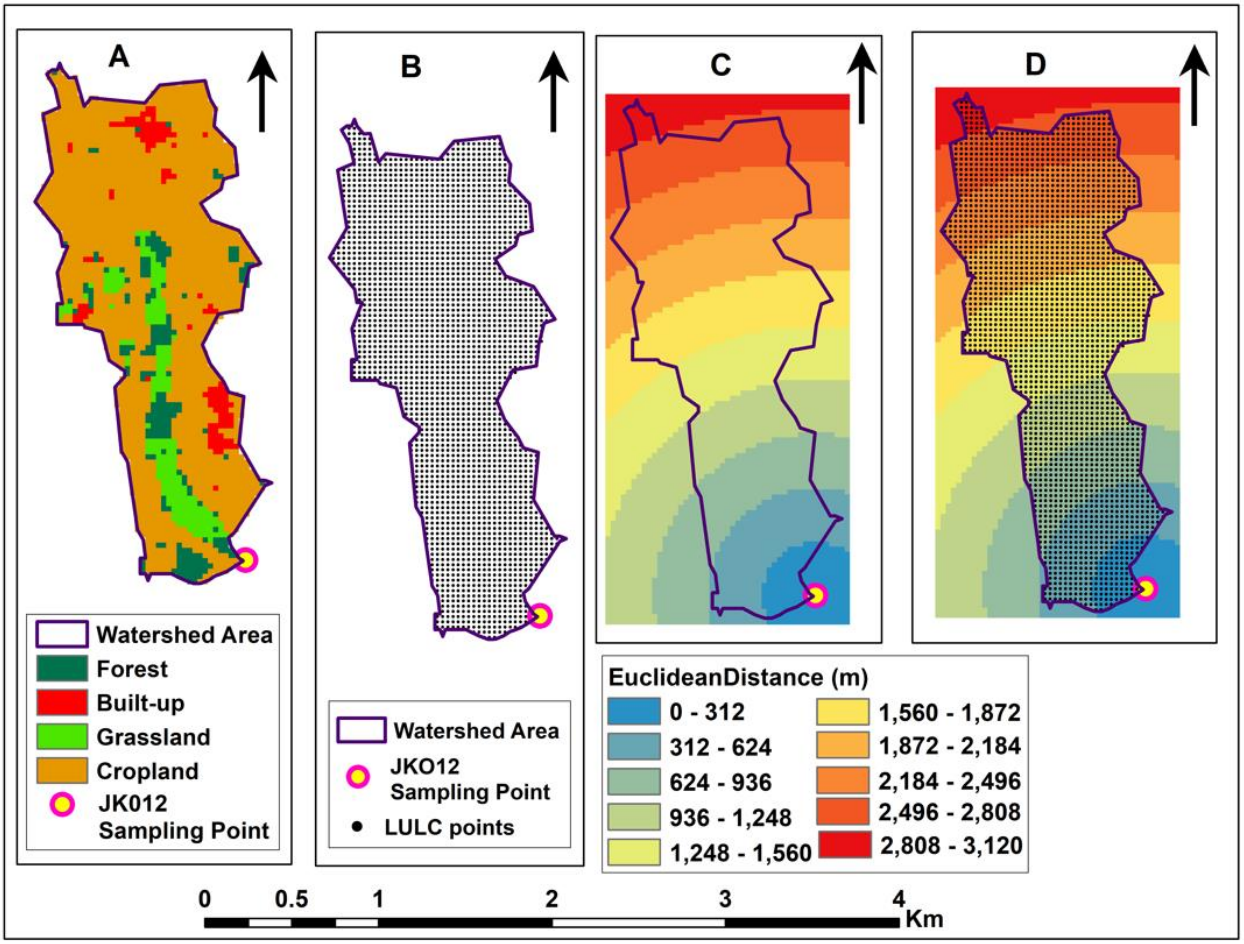


Figure 20: Steps for determining Euclidean distance: (A) Landsat 8 2020 LULC, (B) point features, (C) Euclidean distance raster, and (D) overlaid LULC points layer on Euclidean distance Raster.

4.2.3.6 Weighted Distance

Weighted distance was calculated using WIDF (equation 10) (Peterson *et al.*, 2011).

$$\mathbf{Wi = (d + 1)^{-1}}\text{.....10}$$

Where:

$\mathbf{W_i}$ is the weight of pixel $\mathbf{i}$.

d is the Euclidean distance to the outlet or sampling point.

4.2.3.7 Effective Contribution Area (Aec)

For a specific water quality parameter, the Aec of a LULC class **k** to a designated in-river sampling point is calculated as the product of a unit pixel area and the sum of all **n** upstream pixels belonging to class **k**, each weighted by their distance from the sampling point (Liberoff *et al.*, 2019). This is represented by equation 11:

$$\mathbf{Aec\ (k)} = \boldsymbol{a} * \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Wi\ (k)}\text{.....11}$$

Where:

a represents the unit pixel area, which in this context is (30m * 30m = 900m²).

$\mathbf{W_i}$ denotes the weight assigned to each pixel **i** based on its distance from the sampling point, in this case $\mathbf{W_i}$ is obtained from the inverse function (equation 10).

4.2.3.8 Regression Analysis and Metrices used

Multiple regression analysis was employed to determine the relationship between the WQI and individual water quality parameters with the Aec of each sampling point. The metrics used in the analysis includes coefficient of determination, Significance F and P-value as well as regression coefficients. The coefficient of determination, denoted as R-squared (R^2), was used to measure the proportion of the variance in the dependent variable (water quality parameters) that is predictable from the independent variable (LULC type) in a multiple regression model (Chicco, Warrens and Jurman, 2021). A statistically significant F-statistic and low p-value (< 0.05 in this case) were used to indicate if the multiple regression model explained a significant amount of the variation in the dependent variable, and the independent variables collectively had a significant effect (Chittaranjan, 2019). Regression coefficients show how the dependent variable changes when the independent variable increases by one unit, assuming all other variables are constant. The p-value associated with the regression coefficients determined the statistical significance. A specific independent variable (i.e. cropland) had a significant effect on the dependent variable (i.e. water quality parameter) if the coefficient had a low p-value (below 0.05), indicating statistical significance. The regression

coefficients' sign showed whether the relationship between the independent and dependent variables was positive or negative (Kurtz, 2009; Vogt and Johnson, 2015). The summary of the whole research methodology in the assessment of the linkages between Land-Use/Land-Cover and water quality is illustrated in Figure 21.

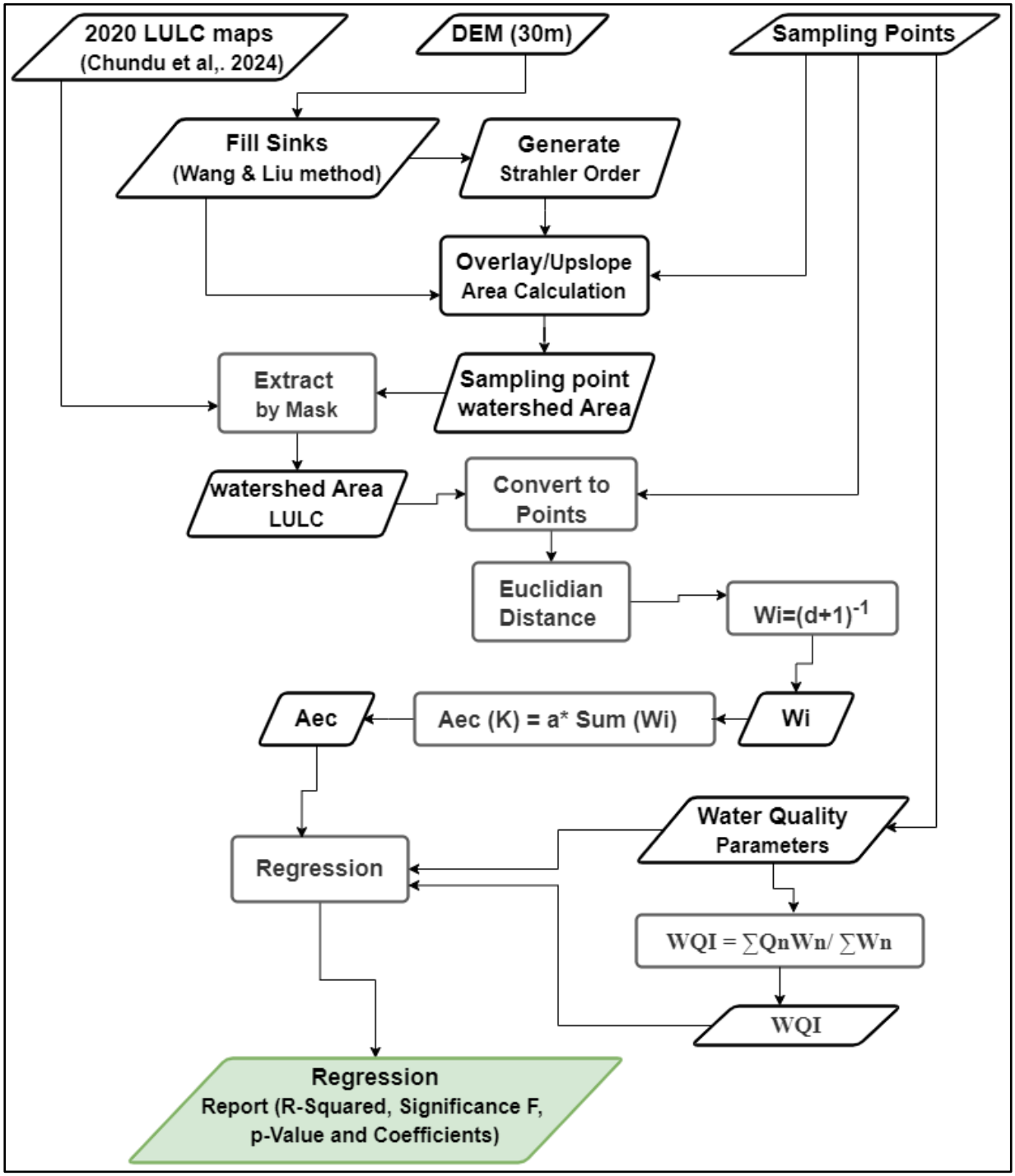


Figure 21: The summary of the research methodology on the assessment of the linkages between Land-Use/Land-Cover and water quality

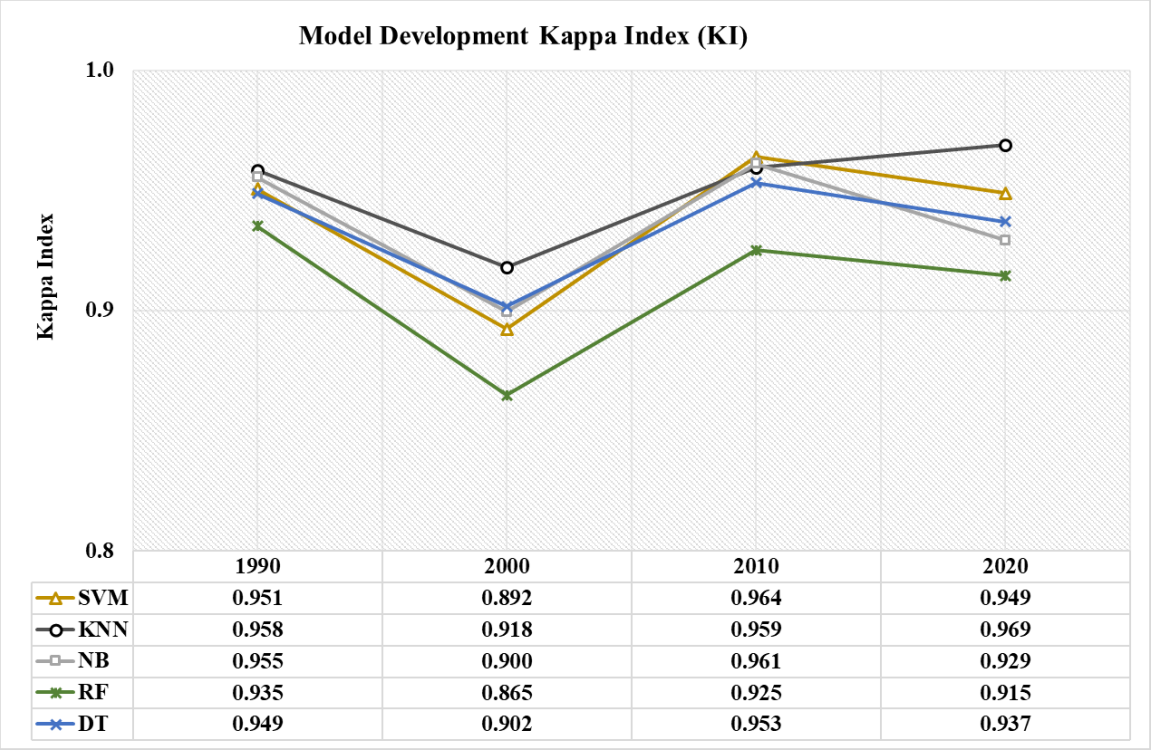
CHAPTER FIVE: RESULTS

This chapter presents the results derived from the data analysis and field findings. This section intends to address the specific objectives and corresponding research questions outlined in the study.

5.1 Specific objective 1: To ensemble a superiorly hybrid machine learning model for enhanced accuracy of modelling LULC changes.

5.1.1 Model Development Accuracy Assessment

The findings indicated that the KNN algorithm exhibited the highest accuracy in model development for the 1990, 2000, and 2020 images, yielding KI values of 0.958, 0.918, and 0.969, respectively. However, for the 2010 images, the SVM algorithm displayed the highest model development accuracy with a KI of 0.964 (refer to Figure 22). Generally, all the algorithms showed consistent and robust performance across different classes of LULC, with F-scores ranging from 0.764 to 1 (Table 9). The highest accuracy was achieved in identifying water bodies, with an F-score of over 99% (Table 9), indicating near-perfect classification results for this specific class. This evaluation suggests that all the algorithms are effective in LULC classification, especially for water bodies, which is critical for various environmental and geospatial applications. ANN model which could not be executed due to a bug in the latest version of QGIS Desktop 3.28.5, therefore, was excluded at this stage and was not considered in the subsequent analysis.



	1990	2000	2010	2020
SVM	0.951	0.892	0.964	0.949
KNN	0.958	0.918	0.959	0.969
NB	0.955	0.900	0.961	0.929
RF	0.935	0.865	0.925	0.915
DT	0.949	0.902	0.953	0.937

Figure 22:Model development Kappa Index of different machine learning models.

Table 9: Model development F-Score of different machine learning models

	Model Development F-Score of Class						Model Development F-Score of Class					
	Model	est	Built-up	Grassland	Water	Cropland		Model	est	Built-up	Grassland	Water	Cropland
	NB	0.969	0.979	0.921	1	0.951		NB	0.912	0.989	0.857	0.998	0.842
	DT	0.964	0.986	0.905	1	0.941		DT	0.911	0.990	0.846	1	0.858
1990	KNN	0.970	0.988	0.920	1	0.953	2000	KNN	0.928	0.990	0.876	1	0.877
	RF	0.959	0.987	0.876	1	0.917		RF	0.871	0.990	0.764	1	0.826
	SVM	0.961	0.988	0.905	1	0.947		SVM	0.912	0.991	0.829	1	0.837

	Model Development F-Score of Class						Model Development F-Score of Class					
	Model	est	Built-up	Grassland	Water	Cropland		Model	est	Built-up	Grassland	Water	Cropland
	NB	0.978	0.980	0.949	1	0.937		NB	0.982	0.978	0.898	0.995	0.865
	DT	0.976	0.986	0.931	1	0.921		DT	0.994	0.985	0.891	0.996	0.883
2010	KNN	0.976	0.989	0.940	1	0.931	2020	KNN	0.984	0.990	0.959	1	0.942
	RF	0.966	0.988	0.868	1	0.880		RF	0.995	0.986	0.831	0.999	0.844
	SVM	0.975	0.988	0.948	1	0.944		SVM	0.996	0.986	0.912	0.995	0.907

5.1.2 LULC classification and triple cross-validation (Post classification accuracy)

The selected models (SVM, NB, DT, RF, and KNN) were used to create LULC maps for the study area. To evaluate the accuracy of the classified maps, a triple cross-validation approach was utilised, incorporating 80 samples organised in triplets (refer to Figure 14 above). Upon assessing the results, it was observed that all models generally exhibited poor performance in classifying cropland in 1990 and 2000, indicated by F-scores ranging from 0.06 to 0.42 (Table 10). This poor performance was attributed to the low quality of the 1990 and 2000 Landsat 5 images which had sparkles. The RF model performed well on the training data but it could not generalise well to the new and unseen data model for the classification of 1990 and 2000 images where it failed to meet the KI threshold of 0.60 (refer to Figure 23 and Table 10). The primary factor contributing to the RF model's low performance was its misclassification of the built-up, grassland, and cropland classes of the 1990 image, with corresponding F-scores of 0.070, 0.011, and 0.264, respectively (refer to Table 10). Furthermore, the model struggled in classifying forest cover and cropland in 2000 images, achieving F-scores of 0.529 and 0.281, respectively (refer to Table 10). The observed low F-score by the RF model in built-up, cropland, and grassland classification could potentially be attributed to the study area's unique characteristics. Specifically, the majority of the study area consists of remote regions characterised by thatched mud houses, which are surrounded by agricultural fields, bare land, fallow land, and grassland. These LULC types may share similarities in terms of color, texture, and shape, making it challenging to accurately distinguish between cropland, grassland, and built-up areas (Lasanta and Vicente-Serrano, 2012; Wambugu *et al.*, 2021). Consequently, the RF model was excluded at this stage and not considered for further analysis. The SVM, NB, DT, and KNN models exhibited strong performance with KI values ranging from 0.60 to 0.86. As a result, these models were selected for subsequent evaluation (refer to Figure 23 and Table 10).

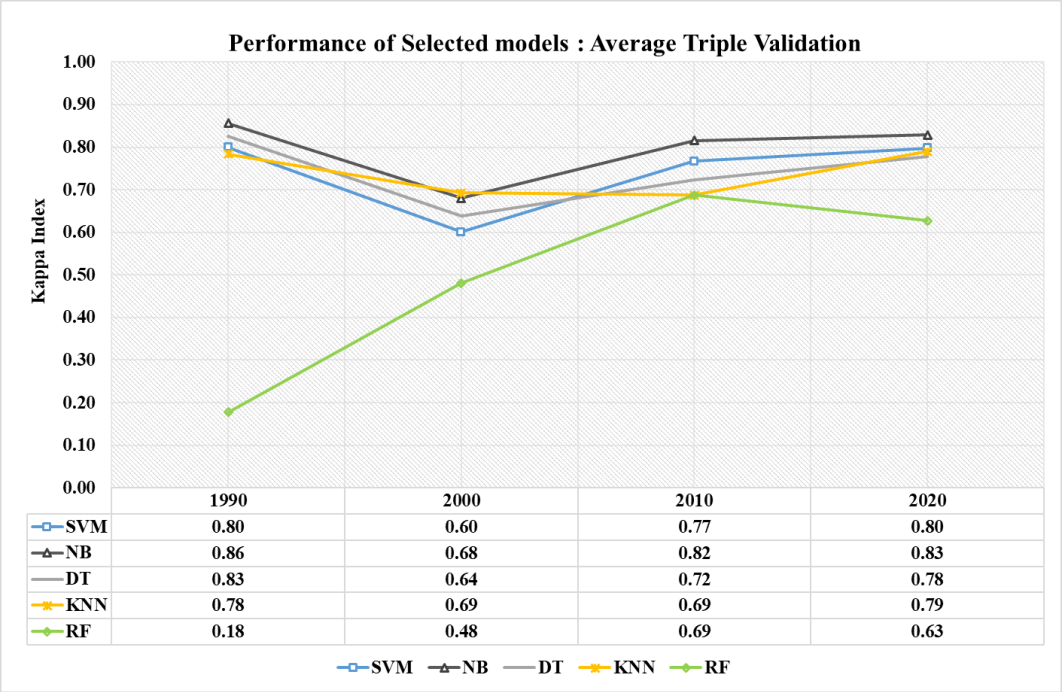


Figure 23: Post-classification accuracy assessment of the performance of different machine learning models.

5.1.3 Fusion of selected models' maps and triple cross-validation

The classified maps generated by the remaining four selected models (SVM, NB, DT, and KNN) were fused using majority voting mechanism in QGIS - Orfeo Toolbox fusion of classifications, to produce a single hybrid map for each of the years 1990, 2000, 2010, and 2020. This fusion of the classified maps from the four models is referred to as the Quad (4) hybrid model. To evaluate the accuracy of the LULC maps derived from the Quad hybrid model, a triple cross-validation approach was employed as well, using the same 80 samples organised in triplets as described earlier (Figure 14 above).

It was observed that the Quad hybrid model developed as a superior machine learning model by combining the classified maps from SVM, NB, DT, and KNN, exhibited superior performance compared to using individual models alone. This was evident from the F-scores, OA, and KI values obtained across all classes (Figure 23 and Table 10). Specifically, for the years 1990, 2000, 2010, and 2020, the Quad hybrid model achieved KI values of 0.87, 0.72, 0.84, and 0.87, respectively (Figure 23). These values indicated a higher level of accuracy and reliability in the classification results compared to using the individual models separately.

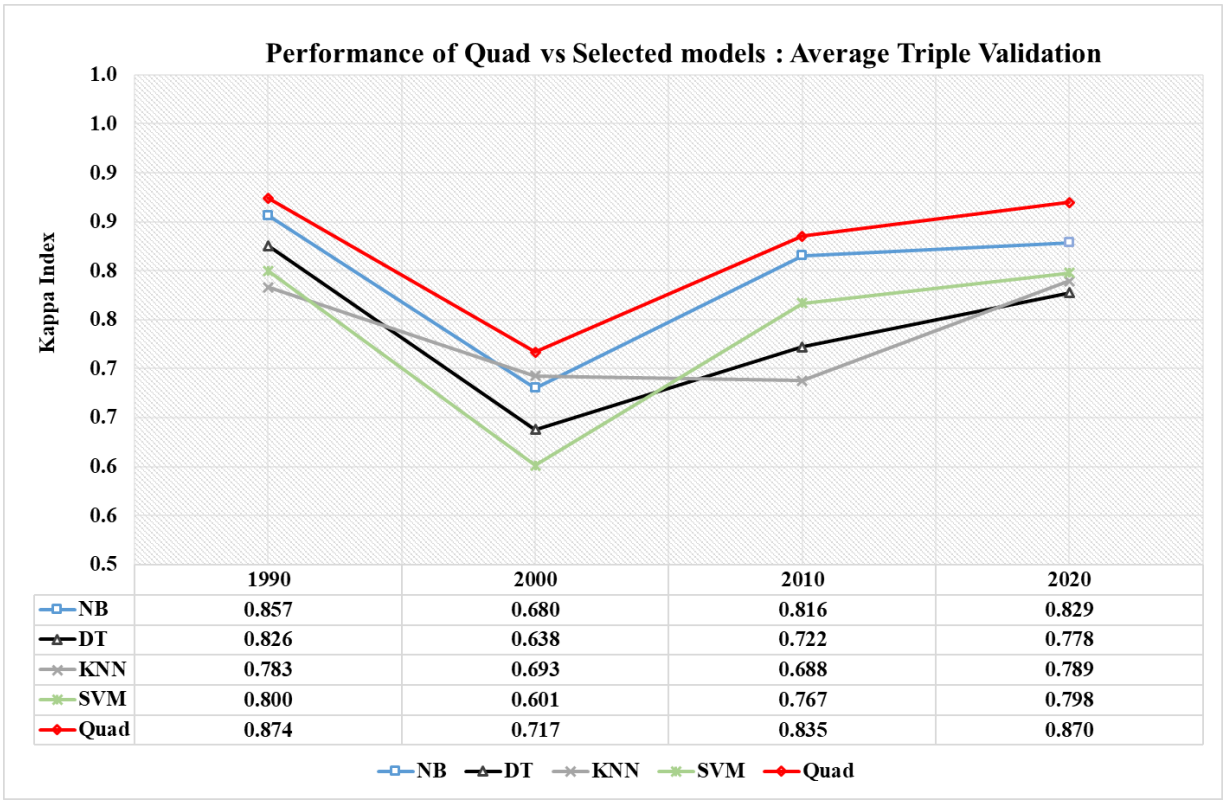


Figure 24: Quad hybrid model performance compared to the four best-selected models

Table 10: F-Scores, Overall Accuracy (OA) and Kappa Index (KI) of different machine learning models

	Average of triple Cross Validation F-Score of Class								Average of triple Cross Validation F-Score of Class							
	Model	Forest	Buil-up	Grassland	Water	Cropland	OA	KI		Forest	Built-up	Grassland	Water	Cropland	OA	KI
1990	NB	0.916	0.952	0.963	1	0.243	0.941	0.857	2000 NB	0.782	0.965	0.919	0.983	0.364	0.876	0.680
	DT	0.884	0.990	0.950	1	0.193	0.924	0.826	DT	0.703	0.991	0.899	1	0.345	0.850	0.638
	RF	0.874	0.070	0.011	1	0.264	0.239	0.177	RF	0.529	0.974	0.803	1	0.281	0.738	0.480
	KNN	0.855	0.966	0.933	1	0.157	0.902	0.783	KNN	0.673	0.980	0.925	1	0.401	0.883	0.693
	SVM	0.735	0.690	0.949	1	0.096	0.913	0.800	SVM	0.781	0.901	0.872	1	0.299	0.818	0.601
	Quad	0.925	0.984	0.962	1	0.329	0.946	0.874	Quad	0.783	0.977	0.925	0.998	0.420	0.888	0.717

	Average of triple Cross Validation F-Score of Class								Average of triple Cross Validation F-Score of Class							
		Forest	Built-up	Grassland	Water	Cropland	OA	KI		Forest	Built-up	Grassland	Water	Cropland	OA	KI
2010	NB	0.712	0.924	0.967	1	0.862	0.943	0.816	2020 NB	0.908	0.895	0.957	0.954	0.480	0.921	0.829
	DT	0.584	0.915	0.938	1	0.675	0.900	0.722	DT	0.800	0.873	0.939	0.962	0.440	0.893	0.778
	RF	0.650	0.965	0.929	1	0.568	0.887	0.688	RF	0.887	0.884	0.852	0.964	0.244	0.793	0.628
	KNN	0.650	0.965	0.929	1	0.568	0.887	0.688	KNN	0.883	0.892	0.942	1	0.418	0.899	0.789
	SVM	0.674	0.924	0.950	1	0.786	0.919	0.767	SVM	0.899	0.877	0.944	0.965	0.440	0.904	0.798
	Quad	0.718	0.969	0.970	1	0.906	0.949	0.835	Quad	0.910	0.897	0.971	0.971	0.567	0.941	0.870

5.1.4 LULC Maps between 1990 and 2020

The Quad hybrid model was employed to produce hybrid maps (from the years 1990, 2000, 2010, and 2020) with five distinct LULC classes, namely forests, built-up areas, grasslands, bodies of water, and cropland. The changes in LULC over the period from 1990 to 2020 were documented and analysed. The findings revealed a general upward trend in land use activities, particularly in the regions encompassing the islands in the wetland and the surrounding wetland areas (Figures 25 and 26, and Table 11).

5.1.5 Bangweulu Wetland and the surrounding area change map from 1990 to 2020

The creation of LULC change maps for the period between 1990 and 2020 involved a comparison of classified images from those respective years. This pixel-by-pixel analysis enabled the identification of areas where changes in LULC had taken place. The resulting change map provided a visual representation of the alterations in the distribution and composition of different LULC classes from 1990 to 2020 (Figures 25A to 25D). The LULC change map (Figure 25E) serves as a valuable tool for comprehending and monitoring environmental changes and evaluating the impacts of human activities. It highlights notable transitions that occurred during the studied period, such as the conversion of more than 13,119 Km^2 of forest cover and 4,821.3 Km^2 of grassland area into cropland (Figure 26). These findings provide valuable insights into the dynamic nature of the landscape and can aid in making informed decisions regarding land use and resource management.

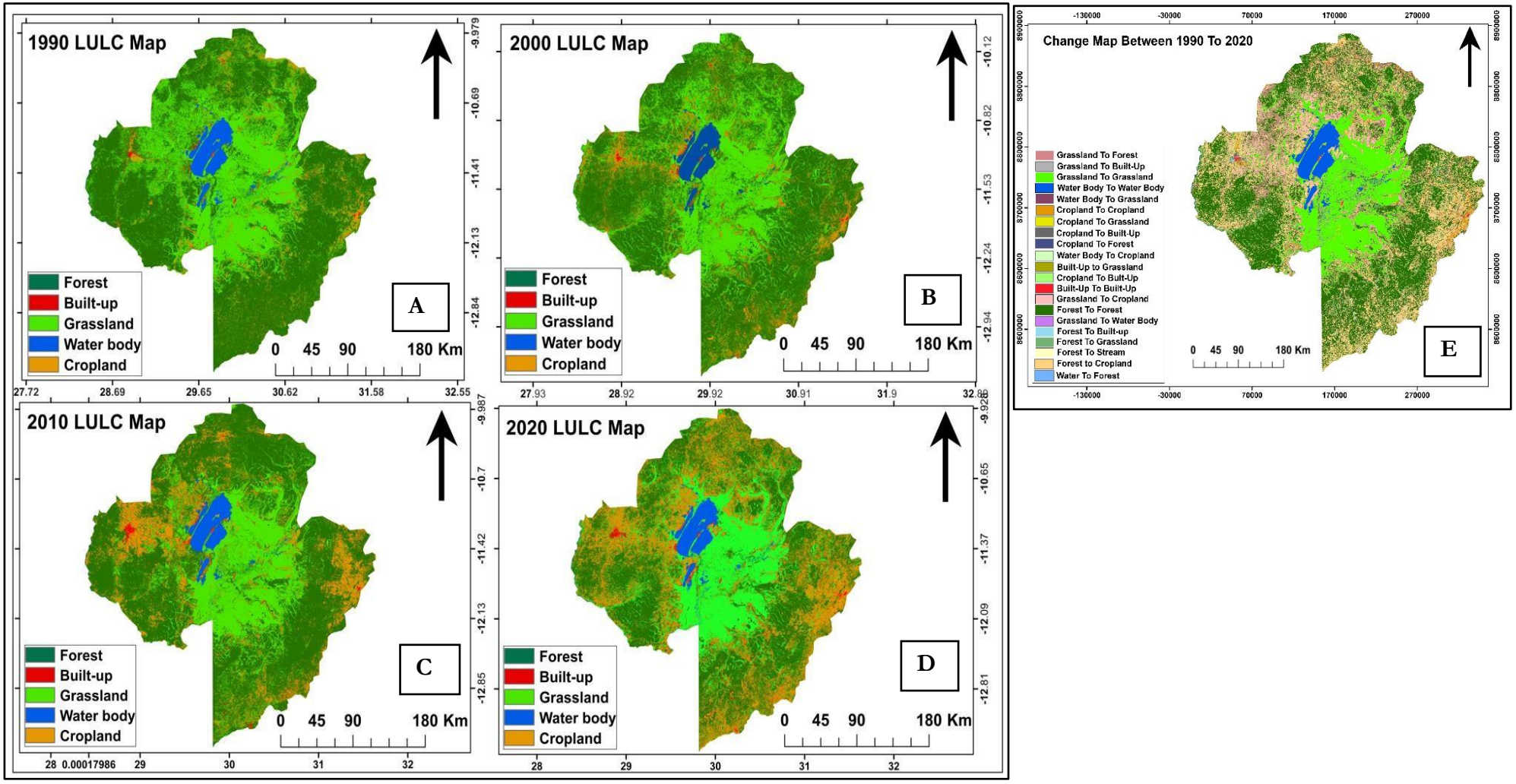


Figure 25: Quad Hybrid Model classified maps of LULC for 1990 (A), 2000 (B), 2010 (C), 2020 (D), and 1990 to 2020 LULC change map (E).

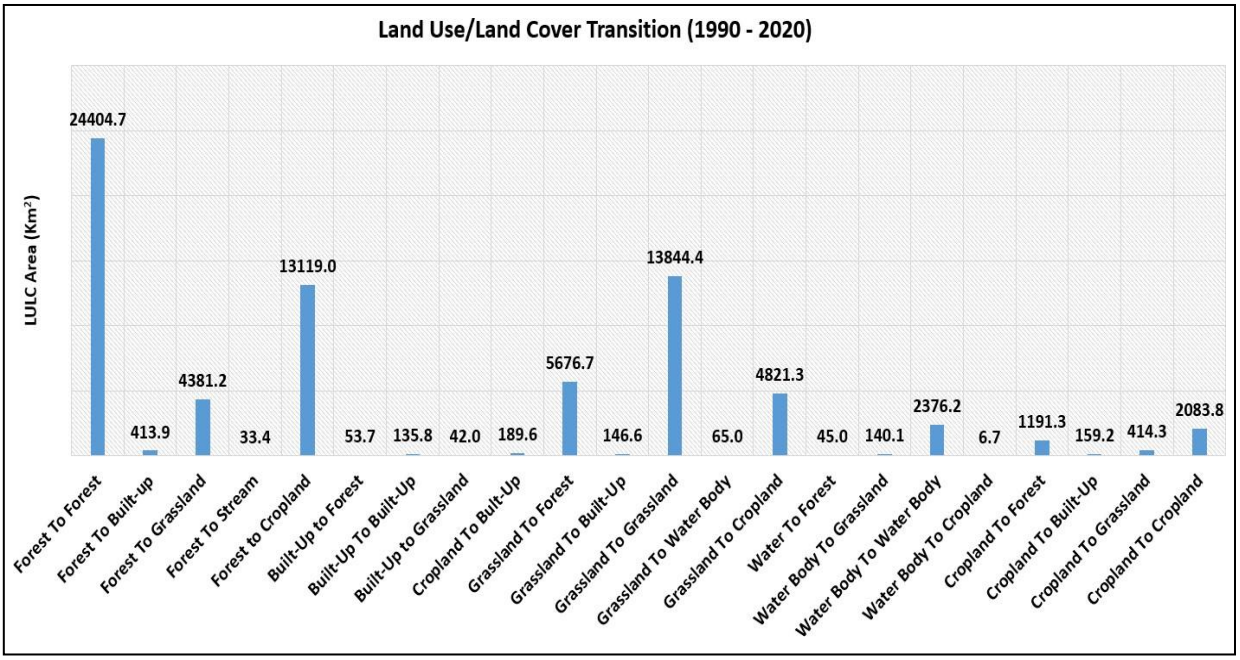


Figure 26: The transition of LULC from one class to another class between 1990 and 2020 in the Bangweulu Wetland and its surrounding areas in Zambia.

The geographical extent and the corresponding percentage coverage values for the LULC maps depicted in Figure 26 are detailed comprehensively in Table 11.

Table 11: Estimates of LULC area and their percent coverage from 1990 to 2020.

Class Name	1990		2000		2010		2020	
	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
Forest	43185.75	57.45	41956.13	55.70	41795.70	55.73	31973.485	42.54
Built-up	428.89	0.571	603.27	0.80	707.44	0.94	873.29	1.16
Grassland	25014.52	33.28	23668.14	31.42	18802.39	25.07	19184.97	25.53
Water	2611.03	3.47	2492.31	3.31	2449.50	3.27	2518.83	3.35
Cropland	3925.11	5.22	6603.23	8.77	11243.94	14.99	20607.57	27.42

5.1.6 Trend Analysis of different LULC classes (1990 – 2020)

The analysis of trend curves for various LULC classes revealed noticeable deviations, encompassing both gains and losses. A comprehensive summary of the LULC variations within the study area from 1990 to 2020, can be found in Table 12, and Figures 27 and 28. Generally, the findings indicated a decline in forest cover, water bodies, and grasslands, while there was an increase in cropland and built-up areas between 1990 and 2020. Forest cover remained relatively stable between 1990 and 2010 but

experienced a significant decrease thereafter until 2020. The grassland exhibited a sharp reduction in size between 1990 and 2010, followed by a period of stabilization. Water bodies also demonstrated a notable decrease from 1990 to 2010, but a positive change was observed between 2010 and 2020 as shown in Figures 27 and 28.

5.1.7 Built-Up Area Change Analysis

The built-up area in the Bangweulu Wetland and its surrounding areas has undergone a significant increase between 1990 and 2020, as observed from the assessment of changes in the classified maps. Specifically, the built-up area has expanded from 428.89 Km^2 to 873.29 Km2, representing a growth of over 50% (Figures 27 and 28, and Tables 12). The period from 1990 to 2000 witnessed the most substantial rise in the built-up area, with an increase of more than 40%, this kind of trend was observed by other studies (Kafy *et al.*, 2020; Taiwo *et al.*, 2023). This was followed by the period from 2010 to 2020, which recorded a growth of 23.44%. The period from 2000 to 2010 experienced a comparatively lower increase of 17.27% (Figure 28 and Table 12). The annual rate of change (ARC) of the built-up areas from 1990 to 2020 was 1.7% (Table 12).

5.1.8 Forest Cover Change Analysis

The forest cover in the Bangweulu Wetland and its surrounding areas exhibited a significant decline between 1990 and 2020. The average ARC for this period was calculated at -1.17%, (Table 12) indicating a notable decrease in forest cover over time. This type of trend was also observed by other studies (Kafy *et al.*, 2022; Rahaman *et al.*, 2022). Specifically, the forest cover decreased from 43,185.75 Km^2 to 31,973.485 Km^2, which corresponds to a decline of 35.07% between 1990 and 2020 (figures 27 and 28, and Table 12). The most substantial decrease in forest cover occurred during the decade from 2010 to 2020, with a reduction of 23.50%. This was followed by the period from 1990 to 2000, which witnessed a decrease of 2.85%. Comparatively, the smallest decrease in forest cover was observed during the period from 2000 to 2010, with a decline of only 0.38% (Figure 28 and Table 12).

5.1.9 Cropland Area Change Analysis

The cropland areas in the Bangweulu Wetland and its surrounding areas exhibited a remarkable increase from 3,925.11 Km^2 to 20,607.57 Km^2, representing a growth of 80.95% between 1990 and 2020 (Figures 27 and 28, and Table 12). A similar trend was also observed by Gxokwe, Dube and Mazvimavi (2023). The average ARC for this period was calculated at 2.7% (Table 12), indicating a consistent upward trend in cropland expansion. The most substantial increase in the area under crop production occurred between 2010 and 2020, with a notable rise of 83.28%. This was followed by a 70.28%

increase between 2000 and 2010 and a 68.23% increase between 1990 and 2000 (Figure 28 and Table 12).

5.1.10 Water Body Change Analysis

This particular LULC class holds significant importance within the study area. Compared to other LULC classes, the observed changes in this class were found to be minimal as also observed by (Kafy *et al.*, 2021; Gxokwe, Dube and Mazvimavi, 2023). Specifically, the water body exhibited an overall decrease from 2,611.03 Km^2 to 2,518.83 Km^2, resulting in a decline of 3.66% between 1990 and 2020 (figures 27 and 28, and Table 12). The average ARC for this period was calculated at -0.12% (Table 12), indicating a slight downward trend in water body coverage. During the period from 1990 to 2000, the water body experienced the highest decrease, with a reduction of 118.72 Km^2 (-4.55%). This was followed by the period between 2000 and 2010, which had a decrease of 1.72%. However, the period from 2010 to 2020 witnessed an increase in the water body, with a growth of 69.33 Km^2 (2.83%) (Figure 28 and Table 12).

5.1.11 Grassland Change Analysis

Between 1990 and 2020, there was a general decrease in the grassland area within the study area, from 25,014.52 Km^2 to 19,184.97 Km^2, representing a decline of 30.39% (refer to Figures 27 and 28, and Table 12). This form of trend was also observed by other studies (Rahaman *et al.*, 2022; Gxokwe et al., 2023). The average ARC for this period was calculated at -1.01 (Table 12), indicating a consistent downward trend in grassland coverage. The period from 2000 to 2010 exhibited the highest percentage decline in vegetation, with a decrease of approximately 20.56%. This was followed by the period from 1990 to 2000, which recorded a decline of 5.38% in the grassland area. However, the period between 2010 and 2020 showed a minimal increase in the grassland area, with a growth of 2.03% (Figure 28 and Table 12).

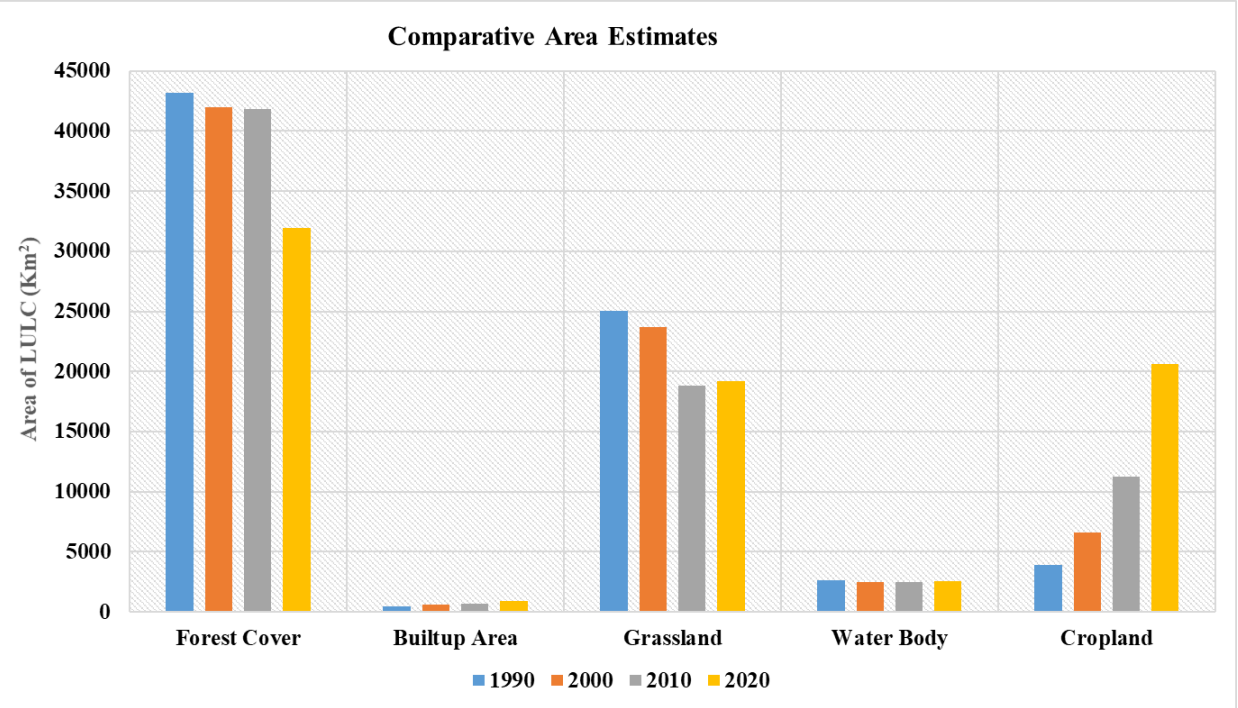


Figure 27: Comparative area estimates for various LULC classes between 1990 and 2020

Table 12: Estimated percentage change in LULC from 1990 to 2020.

	1990 - 2000		2000 - 2010		2010 - 2020		1990 - 2020		
Class name	**Area change (Km^2)**	**% Relative change**	**Area change (Km^2)**	**% Relative change**	**Area change (Km^2)**	**% Relative change**	**Area change (Km^2)**	**% Relative Change**	**% ARC**
Forest Cover	-1229.62	-2.85	-160.43	-0.38	-9822.22	-23.50	-11212.261	-35.07	-1.17
Built-up Area	174.38	40.659	104.18	17.27	165.84	23.44	444.40	50.89	1.70
Grassland	-1346.38	-5.38	-4865.75	-20.56	382.58	2.03	-5829.55	-30.39	-1.01
Water Body	-118.72	-4.55	-42.81	-1.72	69.33	2.83	-92.20	-3.66	-0.12
Cropland	2678.12	68.23	4640.72	70.28	9363.62	83.28	16682.46	80.95	2.70

% Relative change is the difference between the LULC_end and the LULC_start divided by the LULC_start and multiplied by 100 {(LULC_end - LULC_start) / (LULC_start) * 100} %ARC: Percent Annual Rate of Change as described in equation 1 above. (% Relative Change/Time period).

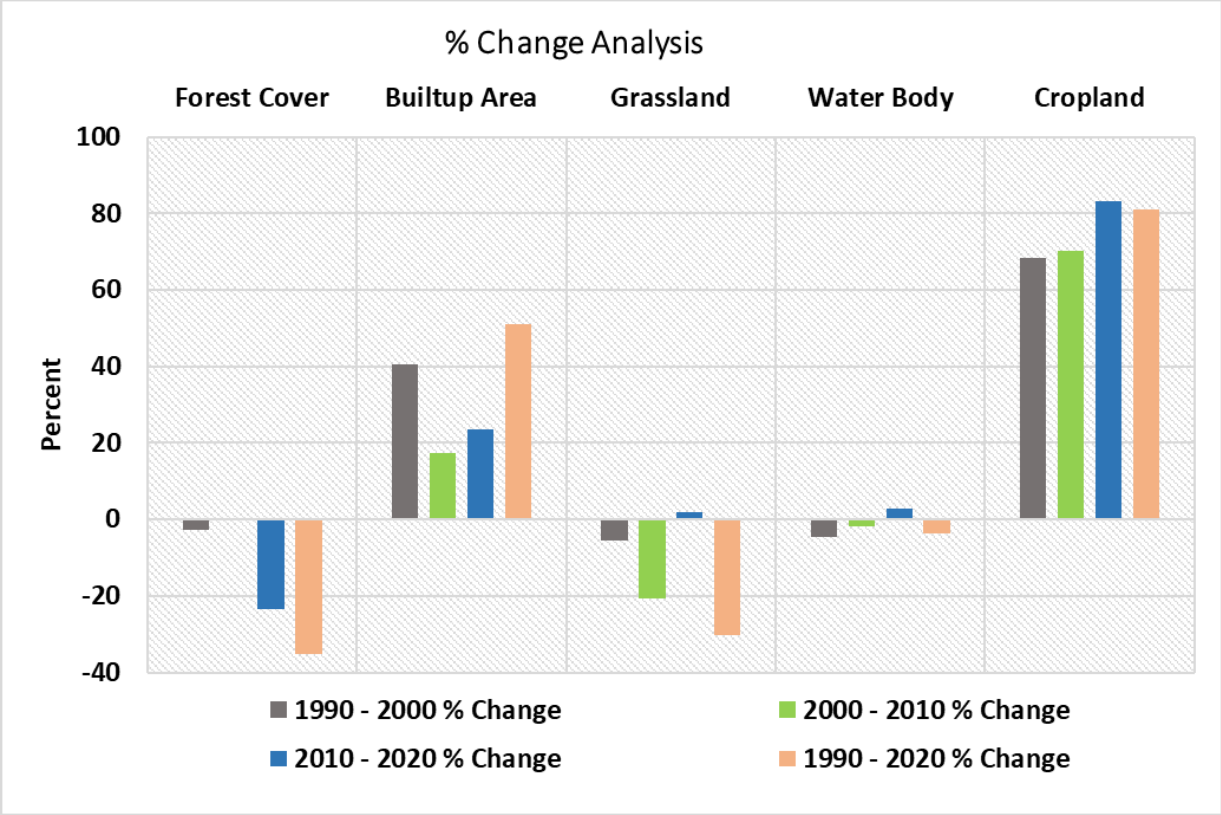


Figure 28: Estimated percent change of different LULC classes between 1990 and 2020

Appendix V number 1 contains the paper published in the Journal of Environmental Challenges based on the results of this specific objective: to develop a superior hybrid machine learning model for improved accuracy in modelling LULC changes.

5.2 Specific objective 2: To rapidly evaluate the variability of water quality.

5.2.1 Modified Normalised Difference Water Index (MNDWI) Map

It was found that the water body's MNDWI values for this study area ranged from 0 to 1 (Figure 29A). The MNDWI values typically vary from -1 to +1, and water bodies are usually represented by values larger than 0, whereas values below zero denote dense vegetation or built-up areas (Adhikari, 2019; Xu, 2006). However, the precise value range for MNDWI water bodies can change based on the research region and the index's particular use (Acharya *et al.*, 2017). For example, some studies have used 0 as the threshold, while others have used different threshold values, such as 0.1, 0.2, and 0.4 (Du *et al.*, 2016; Szabó, Gácsi and Balázs, 2016; Laonamsai *et al.*, 2023). The observed water body extent was then converted to shapefile, which was used for extraction of the boundaries of the various water indices for subsequent analysis.

5.2.2 Normalised Difference Turbidity Index (NDTI)

The study area's water bodies' NDTI values (Figure 29B) ranging from -0.363 to 0.250 show various levels of turbidity. The index ranges between **-1 to +1**. Positive NDTI values imply higher levels of turbidity, whereas negative values indicate clean water (Lizcano-Sandoval *et al.*, 2022). The mean zonal values were extracted from the NDTI map using the geometry points' buffer of 10m of water sampling points for the subsequent analysis as earlier alluded to in section 3.2.2.2b.

5.2.3 Normalized Difference Salinity Index (NDSI) Map

The Normalised Difference Salinity Index (NDSI) has a range of **-1 to +1**, where a value near +1 denotes a high salinity level and a value near -1 a low salinity level (Gerardo and de Lima, 2022). The NDSI map had a value range of -0.318 to 0.733 (Figure 29C), the observed NDSI map indicates that there are variations in the salinity of the water in the study area, with some areas having relatively lower salinity and others having higher salinity. The mean zonal values were extracted from the NDSI map using the geometry points' buffer of 10m of water sampling points for the subsequent analysis (refer to section 3.2.2.2b for the extraction process).

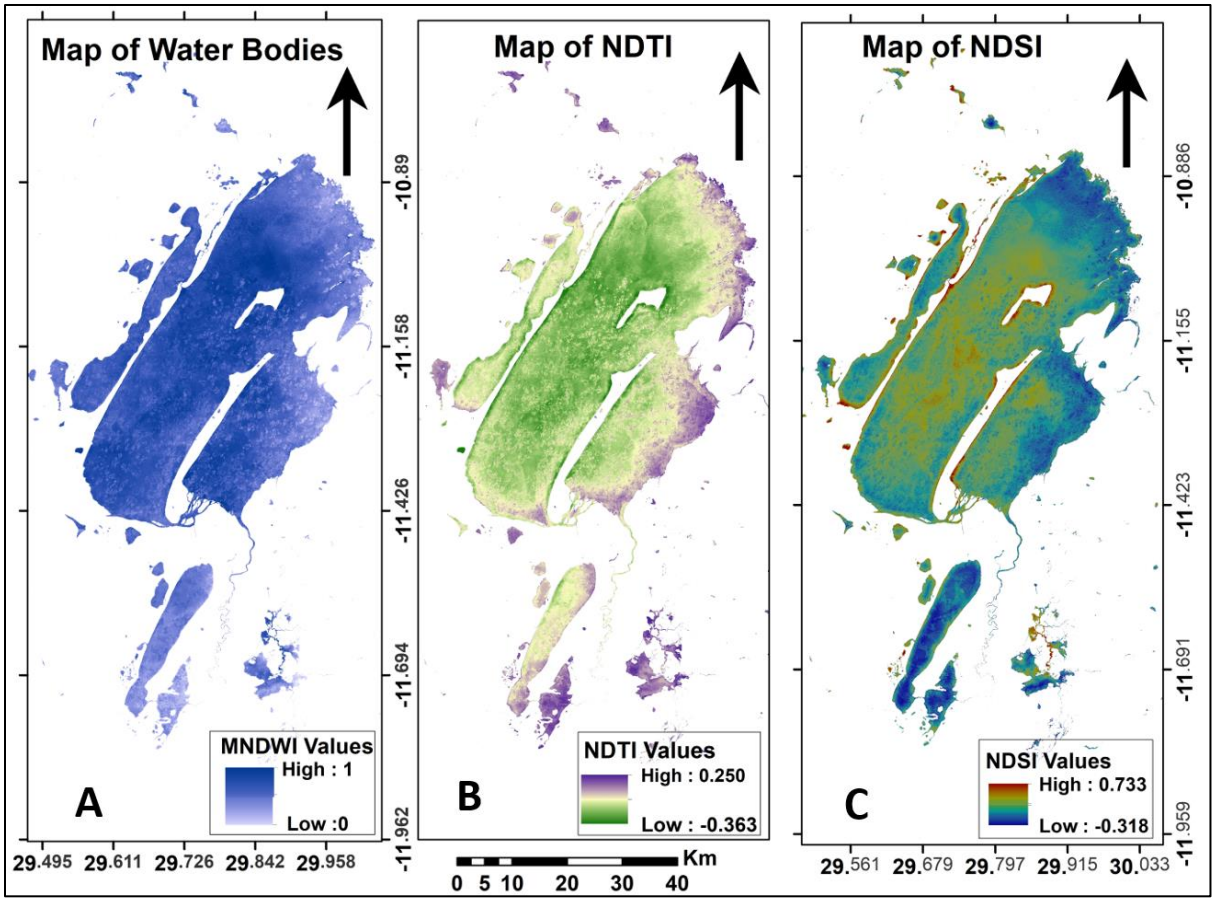


Figure 29: Maps of the Bangweulu Wetland waters (A), normalised difference turbidity index (B), and normalised difference salinity index (C).

5.2.4 Sodium and Chloride maps from the relationship with NDSI

The study used lab-measured concentrations of sodium from 54 field sampling points (34 points for model development and 20 points for validation) to establish the linear relationship between sodium and chloride with the NDSI (using the extracted mean zonal values of NDSI), the R^2 of 0.731 and 0.709 were observed for sodium-NDSI and chloride-NDSI linear relationships, respectively (Figure 30: A and C). The linear regression relationships between chloride-NDSI and sodium-NDSI were used to predict chloride and sodium levels based on 20 laboratory-measured data points, resulting in R^2 values of 0.668 and 0.664, respectively (Figure 31: D and E). The observed linear relationships of NDSI with sodium and chlorides suggest that the index (NDSI) is effective in capturing the presence of sodium and chlorides in water using remote sensing techniques, emphasising the significance of this association in monitoring water salinity levels. Understanding the factors

influencing the relationship of chloride and sodium with NDSI is crucial to account for the percentage that could not be explained by the relationship of these variables.

5.2.5 EC map from the relationship with NDSI

The study used on-site measured EC from 64 field sampling points (44 points for model development and 20 points for validation) to establish a linear relationship between EC and NDSI. The R^2 of 0.734 (Figure 30D), indicates a relatively strong fit model to the data. When the resulting linear regression equation was used to predict the EC using the 20 other points, the predicted EC and the on-site measured EC had an R^2 of 0.718 (Figure 31B).

5.2.6 TDS map from the relationship with EC

The TDS (mg/l) was predicted using 74 field sampling points of the on-site EC measurements through equation 7, where k was 0.7. The predicted TDS was compared with the on-site measured TDS for validation of equation 7, and an R^2 of 0.907 was observed (Figure 31A), indicating a very strong linear relationship between TDS and EC.

5.2.7 Turbidity map from the relationship with NDTI

The study used lab-measured water turbidity from 70 field sampling points (42 points for model development and 28 points for validation) to establish a linear relationship between lab-measured turbidity and the NDTI. The observed relationship had an R^2 of 0.615 (Figure 30B). When the linear regression equation was used to predict the water turbidity, the laboratory measured turbidity and the predicted turbidity had an R^2 of 0.632 (Figure 31C).

5.2.8 Transformation of Index Maps into Water Quality Parameter Maps

The linear equations depicted in Figures 30A, 30C, and 30D, which represent the relationships established between NDSI and measured concentrations of sodium, chloride, and EC, respectively, were used to transform the NDSI map into corresponding sodium, chloride, and EC maps as displayed in Figures 32B, 32C and 32D. Furthermore, empirical equation 7 was used to transform the EC map into a TDS map (Figure 32E). The regression equation between turbidity and NDTI was used to transform the NDTI map into a water turbidity map (Figure 32A).

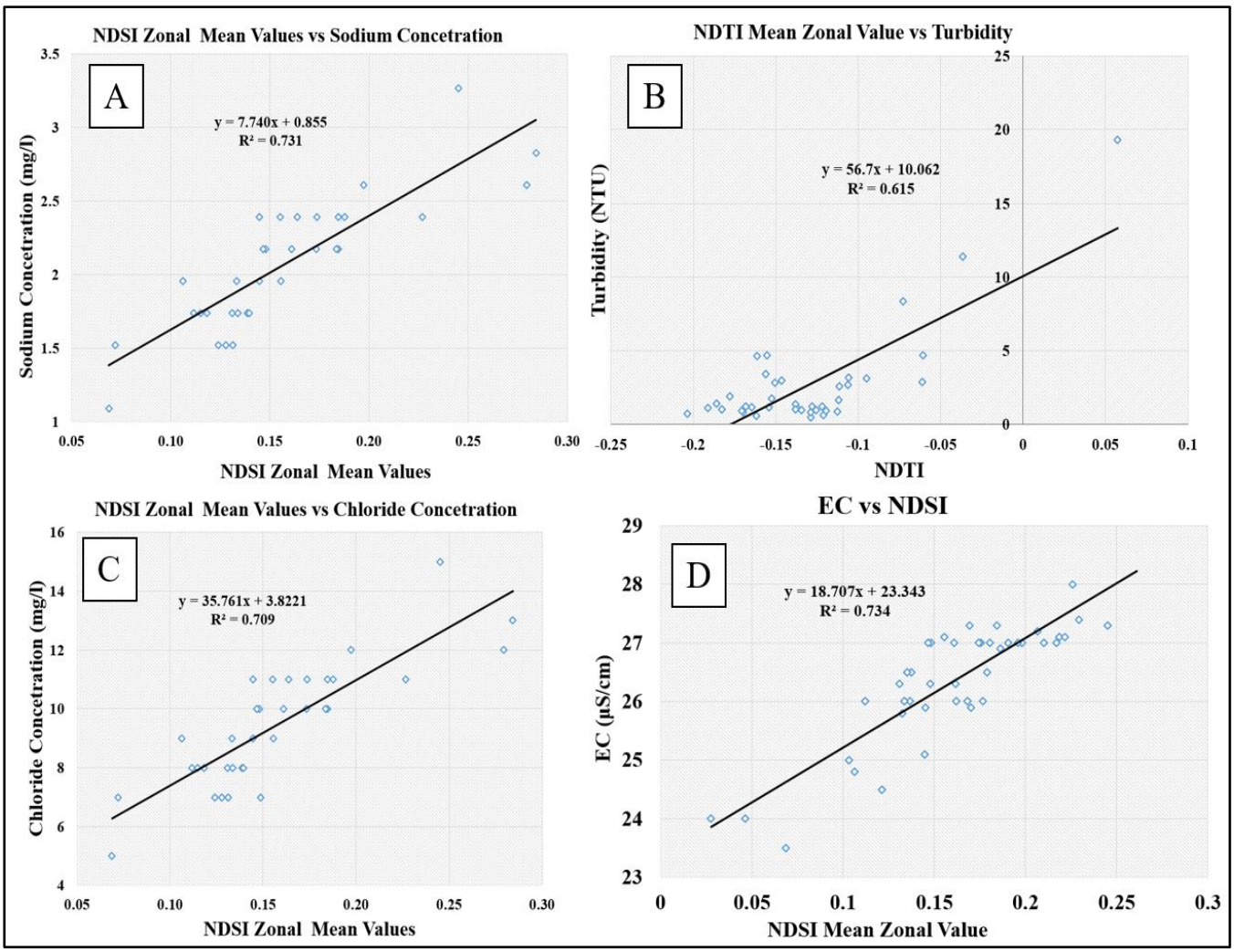


Figure 30: Retrieved relationships between remotely sensed water quality indices and water quality parameters: (A) and (C) relationship of normalised difference salinity index with sodium and chlorides, respectively, (B) relationship between normalised difference turbidity index and turbidity, and (D) relationship between normalised difference salinity index and electrical conductivity.

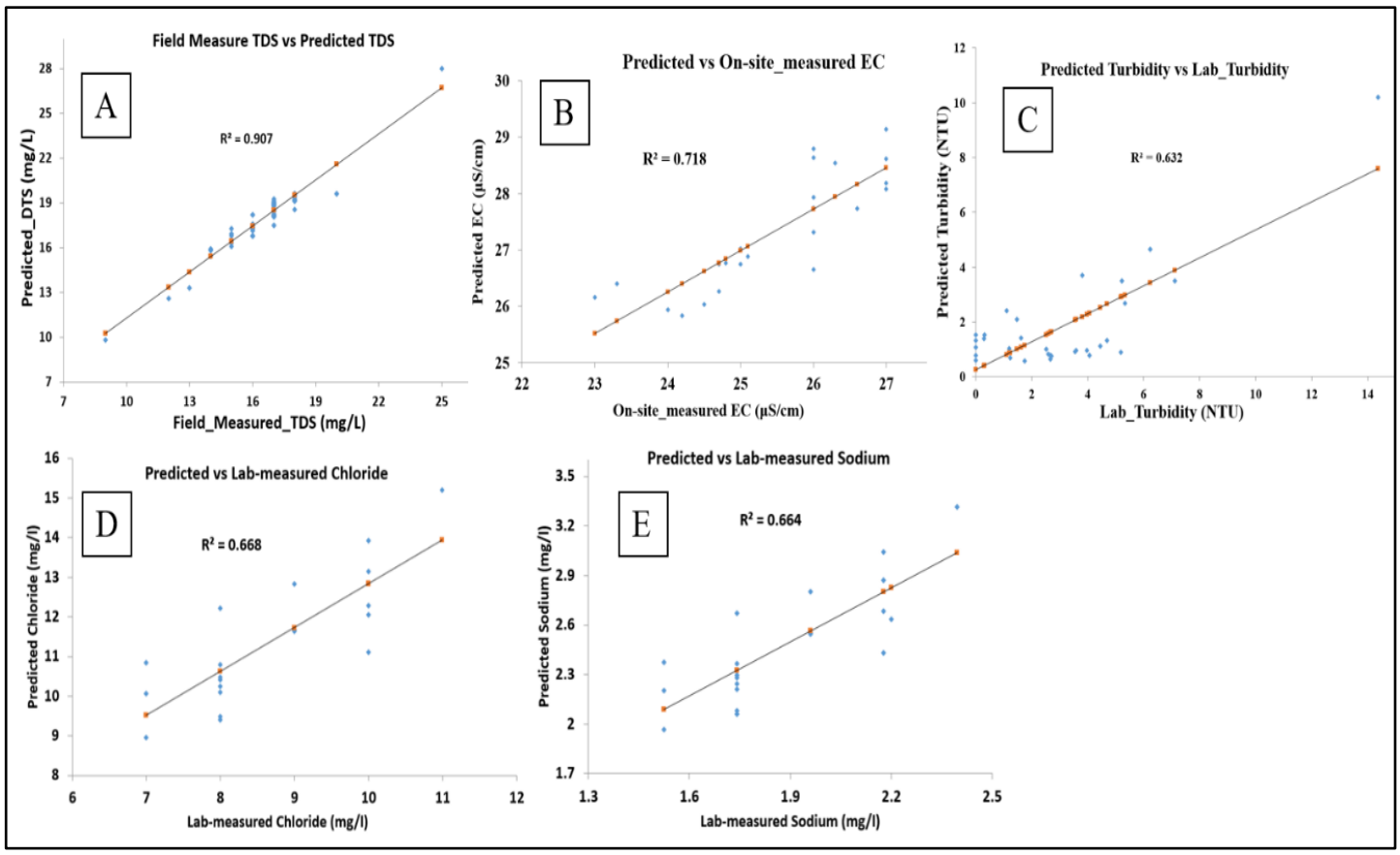


*Figure 31: Validation of the linear equations: (A) relationship between predicted and on-site-measured TDS from TDS=K*EC, (B) relationship between predicted and on-site-measured EC, (C) relationship between predicted and lab-measured turbidity, (D) relationship between predicted and lab measured chloride, and (E) relationship between predicted and lab measured sodium.*

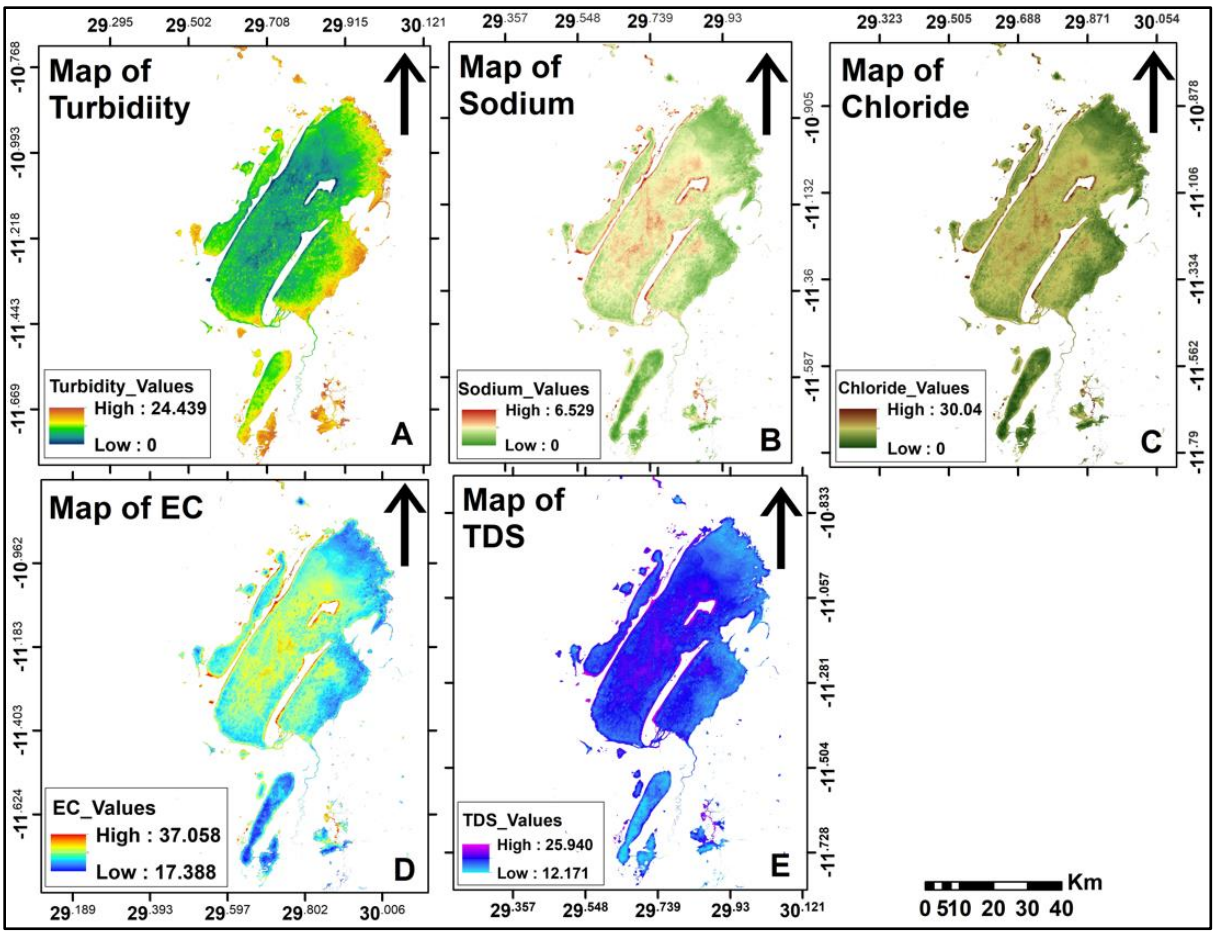


*Figure 32: Maps of water quality parameters: (A) turbidity map from normalised difference turbidity index, (B) sodium, (C) chloride and (D) electrical conductivity maps from normalised difference salinity index, and (E) TDS map from TDS = K*EC.*

5.2.9 Water Quality Index (WQI)

The on-site/lab measurements for seventeen water quality parameters are detailed in Table 7 above. Lead, manganese, nitrates, phosphate, and sulphates were excluded from the Water Quality Index (WQI) calculation due to their concentrations falling below detection limits. Potassium was not included also because it did not have the ambient standard value. The WQI value was derived from the analysis of the mean concentration values of the remaining eleven water quality parameters, as outlined in Table 13. The observed mean WQI value was 37.860, falling within the 'Good' category according to the water quality status classification (refer to Table 8 above).

From the eleven water quality parameters listed in Table 13, only five parameters could be estimated with reasonable accuracy using remote sensing (Figures 30: A to D and 31: A to E). These parameters include EC, TDS, turbidity, sodium, and chloride. Based on these parameters, two Water Quality Indices were calculated (Refer to Table 14). One was derived from on-site/lab data (WQI = 34.948), and the other from remotely sensed (RS) data (WQI = 40.633). The RS-based WQI was obtained by averaging the mean zonal values of the individual water quality parameters at their respective geometry points (as detailed in Table 14). Both WQIs fall within the 'Good' category according to the water quality status classification, implying that the water is suitable for various purposes as described in Table 8 above.

Table 13: WQI calculated from 11 water quality parameters

Parameter	**STD_Value**	**Mean.Conc**	**WQI**
EC (µS/cm)	100	26.086	
TDS (mg/l)	45	16.730	
Turbidity (TNU)	10	2.365	
Sodium (mg/l)	5	1.992	
Chloride (mg/l)	30	9.145	
Faecal (cfu/100ml)	50	28.28	37.860
Calcium hardness(mg/l)	30	7.611	
Total hardness (mg/l)	250	13.440	
Iron (mg/l)	0.7	0.20	
pH	6.5	7.31	
Temp. ^{o}C	23.5	25.630	

STD_Value: Ambient standard values, Mean.Conc: On-site/lab measured mean concentration levels of water quality parameters.

Table 14: Mean on-site/lab-measured water quality parameters and their respective mean remotely sensed water quality parameters.

Parameter	Sampling sites	STD Value	On_site/lab Measured	RS Measured	On_site/lab WQI	RS WQI
EC (µS/cm)	64	100	26.086	26.734		
TDS (mg/l)	74	45	16.730	18.710		
Turbidity(TNU)	70	10	2.365	3.250	34.948	40.633
Sodium (mg/l)	54	5	1.992	2.238		
Chloride (mg/l)	54	30	9.145	10.148		

RS WQI = remotely sensed WQI

5.2.10 WQI map from water quality parameter map

The study generated a total of 180 points, distributed across the entire lakes of the Bangweulu wetland. At each of these points, the mean zonal values (within a 10-meter buffer) of the five (Table 14) remotely sensed water quality parameters were extracted from their respective water quality parameters' maps. The water quality parameter values from various locations served as the basis for calculating the Water Quality Index (WQI) for each location. The calculated WQI from 180 points was used to establish the relationship between WQI and water quality parameters. The observed linear (R^2) relationships between the WQI and water quality parameters were as follows: chloride (0.712), sodium (0.705), TDS (0.705), turbidity (0.628), and EC (0.704) (Figure 33: A to E). These values indicate the strength of the linear relationship between WQI and each water quality parameter and understanding these linear relationships provides insights into how changes in these parameters affect overall water quality. The resulting regression equations (Figure 33: A to E) were then used to transform the water quality parameter maps into WQI maps (Figure 34: A to E).

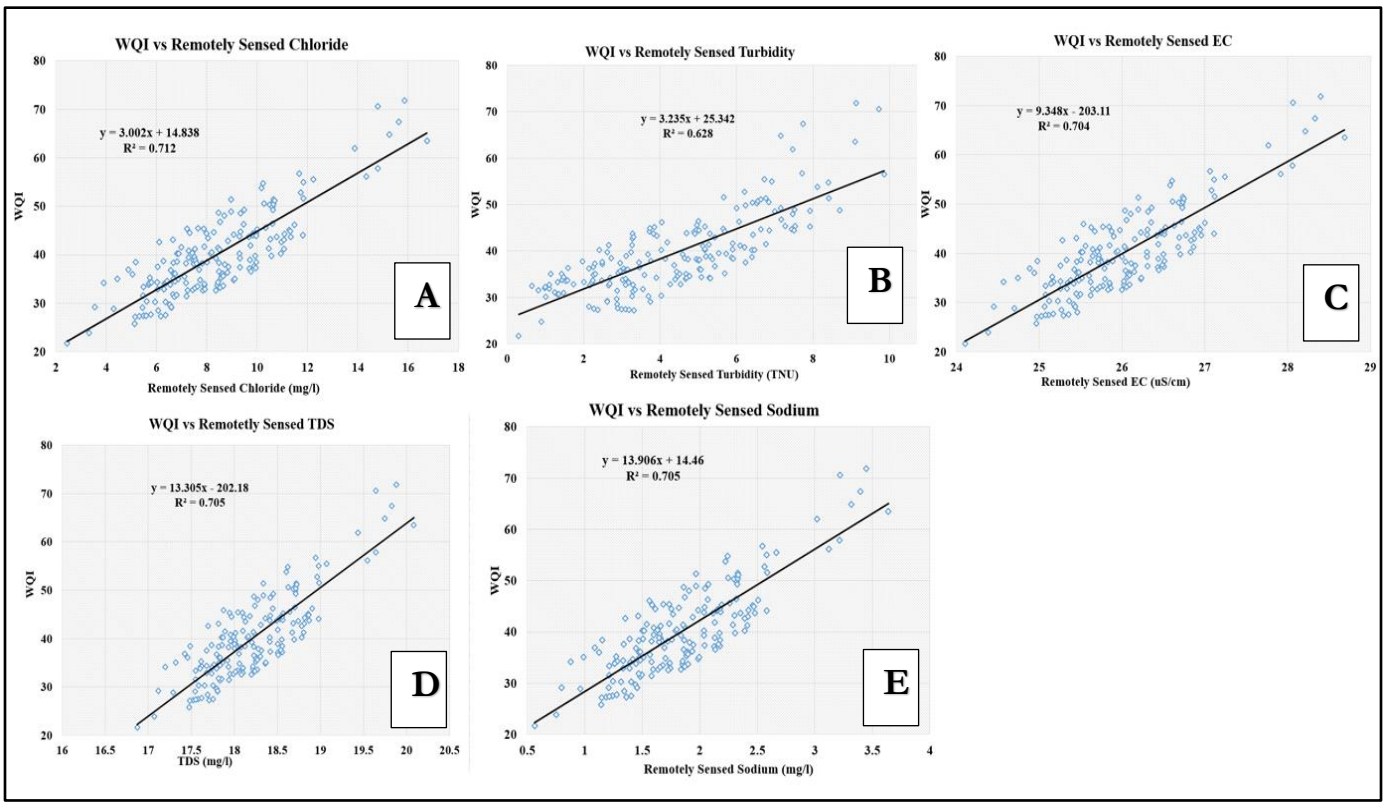
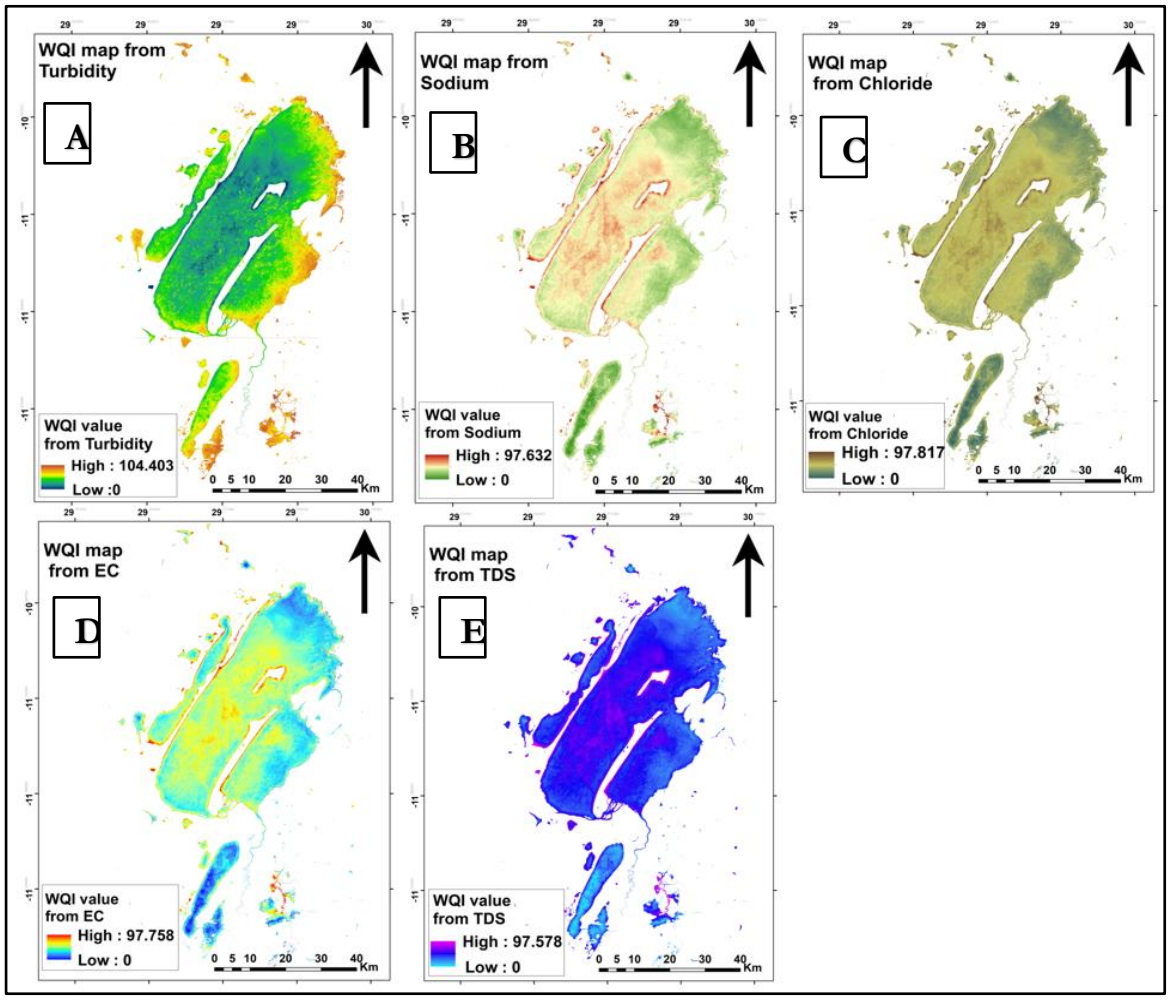


Figure 33: Retrieved relationships between remotely sensed WQI and remotely sensed water quality parameters: (A) relationship between WQI and chlorides, (B) relationship between WQI and turbidity, (C) relationship between WQI and EC, (D) relationship between WQI and TDS, and (E) relationship between WQI and sodium.



5.2.11 Reclassification and Integration of Different WQI Maps

Figure 34: WQI maps: (A) WQI map derived from turbidity map, (B) WQI map derived from sodium map, (C) WQI map derived from chloride map (D) WQI map derived from EC map, and (E) WQI map derived from TDS map.

The manual reclassification procedure was employed to group the Water Quality Index (WQI) maps into distinct classes. This was done to generate meaningful map visualisations, and accurately represent the actual classes within the data, which was based on WQI values ranging from 0 to 25, 26 to 50, 51 to 75, and 76 to 100. The corresponding classes were assigned as excellent, good, poor, and very poor, respectively. The manual approach used in the classification process offered valuable insights into the distribution and variability of water quality parameters, aligning with the water quality status classification approach as outlined in Table 8. To create an integrated WQI map, a weighted overlay approach with equal weights was applied using ArcGIS. This process followed the principle of assigning unit weights (Wn) as suggested by Brown *et al.* (1970). This principle dictates

that the cumulative sum of all assigned weights should be equal to one (∑Wn = 1) as described in equation 9. The specific weights assigned to each of the five water quality parameter maps are illustrated in Table 15, and the resulting maps are depicted in Figure 35: A to F.

Table 15: Assigned weights for different water quality maps

Parameter	Wn	%
EC (µS/cm)	0.2	20
TDS (mg/l)	0.2	20
Turbidity(TNU)	0.2	20
Sodium (mg/l)	0.2	20
Chloride (mg/l)	0.2	20

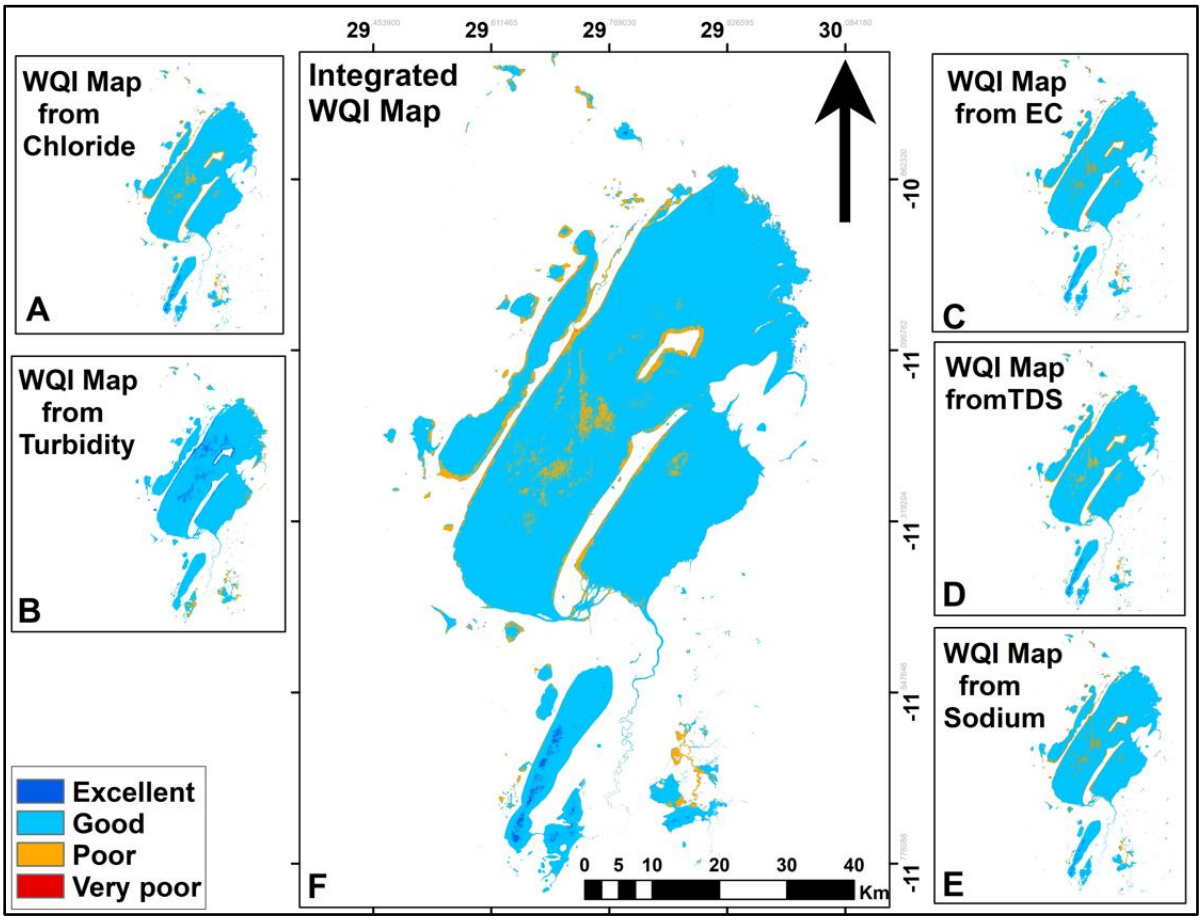


Figure 35: Reclassified WQI maps: A, B, C, D, and E are reclassified maps of WQI from chloride, turbidity, EC, TDS, and sodium, respectively, (F) is an integrated WQI map resulting from the combination of A, B, C, D, and E WQI maps.

Appendix V number 2 contains the paper published in the Journal of Great Lakes Research based on the results of this specific objective: to rapidly evaluate the variability of water quality.

5.3 Specific objective 3: To investigate the influence of LULC on water quality.

5.3.1 In-situ and lab measurement of water quality parameters

Table 16 shows the in-situ and lab measurements of various water quality parameters from 34 sampling points. The presence of iron (Fe^{2+}) in water emerges as a major contributor to elevated WQI values (Figure 36), as seen from various sampling points with higher iron levels like Nkulumashiba Bridge (iron level = 2.23mg/l), Chisangwa 1 (iron level = 1.09mg/l), JK002 (iron level = 2.23mg/l), and JK003 (iron level = 1.09mg/l), where Fe^{2+} levels exceeded the threshold (0.7mg/l, refer to Table 16), resulting in elevated WQI values of 255.543, 148.342, 149.41, and 123.71, respectively (refer to Figure 36). Whereas sampling points like MC18 (Fe^{2+} level = 0.121mg/l) and JK008 (Fe^{2+} level = 0.031mg/l), where Fe^{2+} levels were below the threshold (0.7mg/l), resulted in low WQI values of 35.033 and 32.934, respectively (refer to Figure 36).

Table 16: In-situ and lab measurements of various water quality parameters.

Sampling Point	EC	TDS	TURB	Na^{+}	Cl^{-}	Feacal	Ca^{+2}	TH	Fe^{+2}	Ph
MC 9HM	23	15	18.2	4.62	7	0	17	30	0.23	6.87
MC 12HM	23	15	6.52	7.92	12	>100	17	30	0.23	6.87
MC16	18	12	1.83	36.3	55	20	20	34	0.44	6.82
Mpanta_JK009	28	18	14.8	6.6	10	0	4	28	0.82	7.21
MC20	26	17	2.84	6.66	10	0	14	20	0.23	7.26
MC 1 HM	23	15	6.4	6.6	10	0	17	30	0.23	6.87
MC 8 HM	23	15	10.2	9.9	15	0	17	30	0.23	6.87
MC 11HM	23	15	19.3	6.6	10	0	17	30	0.23	6.87
MC18	25	16	4.24	5.28	8	0	14	20	0.12	6.90
MC17	19	13	8.36	7.92	12	44	14	26	0.18	6.49
Kasongole	26	17	8.36	7.92	12	0	6	26	0.18	6.78
JK009	28	18	14.8	6.6	10	50	10	28	0.24	7.21
Mulembwa	51	34	9.13	9.9	15	0	6	32	0.30	7.38

Sampling Point	EC	TDS	TURB	Na^{+}	Cl^{-}	Feacal	Ca^{+2}	TH	Fe^{+2}	Ph
JK011	36	26	8.82	6.6	10	0	16	40	0.23	7.46
R2-20	27	17	5.1	9.21	14	30	1	38	0.62	8.03
MC 15HM	25	16	34.2	6.57	10	>100	0	30	0.00	6.73
MC15	25	16	34.2	6.57	10	>100	26	30	0.02	6.73
JK008	25	17	7.49	7.89	12	0	16	32	0.03	6.79
MC14	23	9	12.4	5.28	8	0	14	30	0.62	6.68
JK007	51	34	9.13	9.9	15	0	14	32	0.58	7.38
R2-23	27	17	4.7	9.2	14	40	1	38	0.62	7.81
MC19	26	16	1.39	9.9	15	0	20	50	0.03	7.19
Musamfwe	34	24	2.98	8.6	13	0	8	40	0.58	6.17
JK012	40	28	9.23	8.5	13	0	20	28	0.49	6.79
JK005	86	58	6.85	9.9	15	50	10	44	0.30	6.85
Mukuku	27	18	6.34	9.9	15	0	7	20	0.15	6.36
R2-28	40	25	10.41	9.2	14	>100	1	38	0.21	7.15
JK003	67	50	11.2	5.9	9	10	6	60	0.94	7.02
Chisangwa 2	67	50	11.2	5.9	9	0	2	60	0.94	7.02
Mansa River	25	18	6.79	6.6	11	0	6	40	0.21	7.04
JK001	79	55	2.98	8.6	13	0	20	40	0.58	6.73
JK002	90	65	12	9.9	15	10	20	40	1.09	7.18
Chisangwa 1	90	65	12	9.9	15	0	8	40	1.09	7.18
Nkulumashiba	27	20	12.7	6.6	10	0	4	26	2.23	6.36

Abbreviations and units: Total Dissolved Solids-TDS (mg/l), Sodium-Na^{+} (mg/l), Chloride-Cl^{-} (mg/l), Total hardness-TH (mg/l), Iron Fe^{2+}(mg/l), Turbidity-Turb (TNU), Feacal (cfu/100ml), Electrical Conductivity-EC (µS/cm).

5.3.2 Water Quality Index

The variations in WQI values at different water sampling points (34) in the catchment ranged from 32.941 to 255.543. This variability in WQI values among sampling points was also noted by Ahmed et al. (2023) and Wang et al. (2017). These variations indicate differences in water quality across locations, with some points showing better water quality than others based on the calculated WQI values as shown in figure 36.

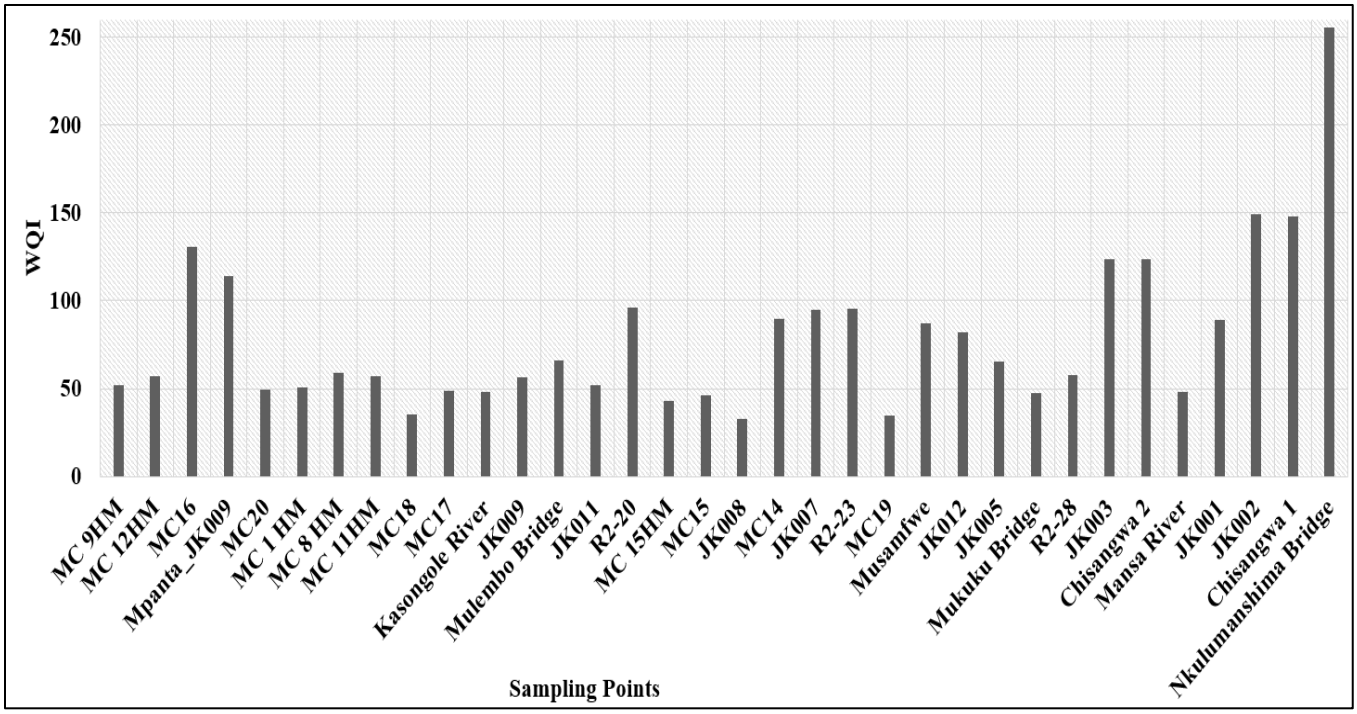


Figure 36: The variations in WQI values across 34 water sampling points.

5.3.3 Effective contribution area

The variations in Aec values for different LULC types (forest, grassland, cropland and built-up) within the catchment varied significantly. The range was from 1 to 5967.726m² for forest, 1 to 1144.025m² for built-up areas, 1 to 1964.221m² for grassland, and 1 to 7098.988m² for cropland (Figure 37). This kind of variations were expected as it was also observed by (Liberoff *et al.*, 2019). These values indicate the significant spatial heterogeneity in LULC patterns across the catchment, reflecting varying degrees of impact and utilisation of different land types within the area under studied.

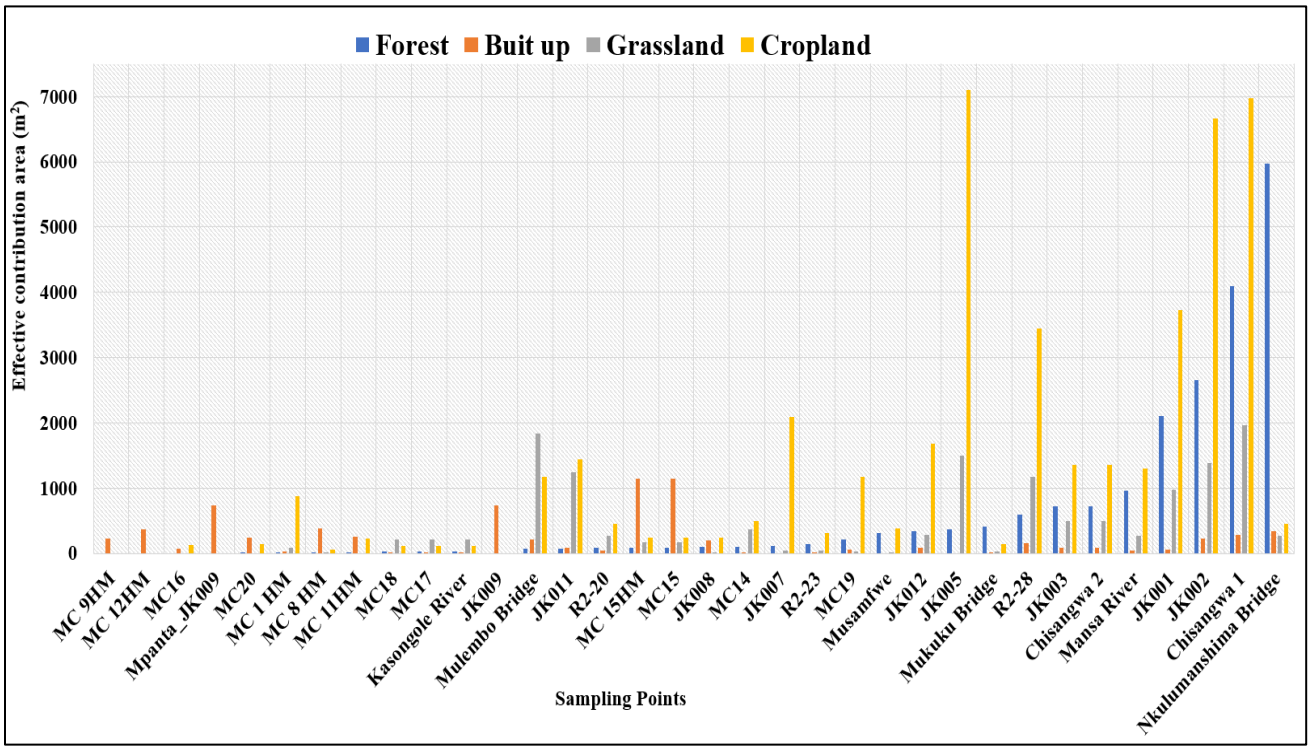


Figure 37: The variations in Effective Contribution Area values for different LULC types

5.3.4 Influence of LULC on water quality

As shown in Table 17, the study utilised multiple regression analysis to determine the relationship between the WQI/individual water quality parameters with the Aec of each sampling point. The results (Table 17) revealed several significant correlations. Turbidity, TDS, Fe^{2+}, and EC showed a strong correlation with LULC, with R^2 values greater than 0.7 and significance F less than 0.05. These parameters had a positive coefficient, indicating a direct positive relationship. The analysis further highlighted the specific relationships between these water quality parameters and different LULC types. Turbidity was more strongly correlated with built-up areas, with a p-value less than 0.05. Fe^{2+}, on the other hand, was more correlated with forested areas, with a p-value of 0.05. TDS and EC were more correlated with cropland, with a p-value less than 0.05. Additionally, feacal coliform was more correlated with built-up areas, with a p-value of 0.006 and a positive coefficient, although it had a relatively lower R^2 of 0.258 and a significance F of 0.063. In contrast, certain water quality parameters, such as K^+, Na^+, Cl^-, and Ca^{2+}, showed a low correlation with LULC, with R^2 values of less than 0.1. Furthermore, the study examined the relationship between LULC and the overall WQI, which demonstrated a reasonable correlation, with an R^2 of 0.649. Appendix V number 3 contains the paper published in the Journal of Physics and Chemistry of the Earth based on the results of this specific objective: to investigate the influence of LULC on water quality.

Table 17: Multiple Regression Summary Output (Confidence level 95%)

Regression

Statistics

	WQI	TH	Fe^{+2}	K^{+}	Na^{+}	Cl^{-}	Ca^{+2}	Turbidity	TDS	pH	EC	Feacal
R Square	0.649	0.199	0.701	0.034	0.033	0.034	0.091	0.708	0.759	0.140	0.799	0.258
Standard Error	28.678	9.102	0.256	1.783	5.406	8.175	7.147	4.647	8.481	0.349	10.346	40.256
Observations	34											

ANOVA

Significance F

	df	WQI	TH	Fe^{+2}	K^{+}	Na^{+}	Cl^{-}	Ca^{+2}	Turbidity	TDS	pH	EC	Feacal
Regression	4	0.000**	0.155	0.000**	0.902	0.910	0.906	0.580	0.000**	0.000**	0.340	0.000**	0.063*
Residual	29												
Total	33												

P-value

	WQI	TH	Fe^{+2}	K^{+}	Na^{+}	Cl^{-}	Ca^{+2}	Turbidity	TDS	pH	EC	Feacal
Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.457
Forest	0.000**	0.672	0.000**	0.770	0.763	0.766	0.356	0.802	0.439	0.113	0.662	0.192
Buit up	0.564	0.388	0.265	0.428	0.443	0.436	0.803	0.000**	0.830	0.963	0.883	0.006*
Grassland	0.711	0.828	0.765	0.714	0.726	0.719	0.183	0.669	0.263	0.223	0.263	0.873
Crop land	0.72	0.213	0.506	0.650	0.657	0.654	0.156	0.956	0.000**	0.989	0.000**	0.499

	Coefficients											
	WQI	**TH**	**Fe+2**	**K+**	**Na^+**	**Cl^-**	**Ca+2**	**Turb**	**TDS**	**pH**	**EC**	**Feacal**
Intercept	67.252	32.748	0.3517	3.067	9.226	14.034	11.89	5.142	16.034	6.9343	24.353	7.816
Forest	0.030	-0.0006	0.0003	-8E-05	-3E-04	-4E-04	-0.001	-2E-04	0.001	-1E-04	0.0007	-0.008
Buit up	-0.010	-0.0047	-2E-04	-8E-04	-0.002	-0.004	0.001	0.0228	-0.001	-1E-05	-9E-04	0.0707
Grassland	-0.006	0.001	-4E-05	-3E-04	-1E-03	-0.002	-0.005	0.001	0.005	0.0002	0.0061	-0.003
Crop land	-0.002	0.0018	-3E-05	1E-04	4E-04	0.0006	0.002	4E-05	0.0055	-8E-07	0.008	0.0043

** highly significant, * significant.

CHAPTER SIX: DISCUSSIONS

This section offers a comprehensive analysis of the results based on specific objectives. It contextualises these findings within the broader framework of the study and examines their implications, limitation as well as uncertainties.

6.1 Assembling of a superiorly hybrid machine learning model for enhanced accuracy of modelling LULC changes.

6.1.1 Performance of different models

As earlier alluded to in section 2.3, it's important to note that the performance of ML algorithms varies depending on the specific dataset, the complexity of the problem, the availability of data, sample size, methods, time and space. As the "No Free Lunch Theorem" suggests: there is no universal ML algorithm that performs best on all possible problems. This theorem highlights the fundamental limitations of ML algorithms in achieving optimal performance across all problem domains (Wolpert and Macready, 1997). Therefore, there are several reasons why the models in the study performed differently from one another as observed in Table 10. First of all, each ML method has its strengths and weaknesses. For example, some algorithms are better at handling complicated interactions and noisy data than others, which could be useful for specific LULC types (Adugna, Xu and Fan, 2022). Secondly, the models' performance might have been affected differently by the dimensionality, noise level, and class imbalance of the input data (Sun, Wong and Kamel, 2009; Chao *et al.*, 2022). Thirdly, there is a possibility that some algorithms were more successful for particular classes because different LULC classes typically exhibit different degrees of separability in the feature space (Abdi, 2020).

The RF model was excluded from further consideration as it did not meet the desired performance expectation (KI > 0.6) for the classification of the 1990 and 2000 images as depicted in Figure 23 above, indicating that it was a weaker model compared to DT, SVM, K-NN, and NB based on the methods used in this research. The RF is an ensemble model that utilises several decision trees as base classifiers, thus it was expected that it would outperform DT. However, in this study, DT performed better than RF which contradicts several studies (Gislason, Benediktsson and Sveinsson, 2006; Kulkarni and Lowe, 2016). This was possible because this research used the minimum training sample size based on the previous literature.

DT can outperform the RF algorithm when the training sample size is limited due to its ability to create more flexible, simpler, and finer-grained decision boundaries which could

potentially capture the specific patterns present in the limited data more effectively (Breiman, 2001a). Thakur and Panse (2022) also observed that it's easier to assess the performance of DT than RF with smaller data sets. Therefore, the underperformance of the RF model could potentially be attributed to the limited size of the training samples used in this research, assuming that the set parameters (Table 5 above) were optimal. In addition, the RF is an ensemble algorithm, which relies on the diversity and quality of weak learners (several decision trees), it typically benefits from larger and more diverse training data sets to create complex and stronger models, and this was also observed by Kulkarni and Lowe (2016) who stated that RF works well when provided with large homogeneous training data.

Thakur and Panse (2022) also added that RF requires a lot of resources and computational power as it relies on several decision trees for image classification. So, with the limited training sample size used in this research, the RF algorithm might have struggled to generalize well to unseen instances, as the limited training data set might not have adequately represented the full variability of the dataset, which could have restricted their ability to leverage the advantages of ensemble learning and the benefits of multiple models thereby creating an unstable decision tree with reduced accuracy (Caruana and Niculescu-Mizil, 2006; Polikar, 2006; Fernández-Delgado *et al.*, 2014).

The KI and OA obtained from the four best performing models (SVM, NB, DT, and KNN) for the LULC classification of the BWS and the surrounding areas (Table 10) were reasonable based on Table 6 and consistent with several other studies (Tehrany, Pradhan and Jebuv, 2014; Basheer *et al.*, 2022; Bayas *et al.*, 2022). However, it is important to note that the performance of these four models exhibited differences in the values of the KI and OA. The two accuracy metrics (KI and OA) varied not only among the models but also with space and time (refer to Table 10). This could be attributed to atmospheric, surface, and illumination fluctuations, data sets, feature representation, hyperparameter adjustment, and the individual issue domain as observed by several other studies (Foody, 2008; Rwanga and Ndambuki, 2017; Leyk *et al.*, 2018; Talukdar *et al.*, 2020).

6.1.2 Hybrid machine learning

The fusion of the four models' maps (SVM, NB, DT, and KNN) produced the Quad hybrid maps with a higher KI and OA (Figure 24 and Table 10) than any of the four models, showing that the Quad hybrid model's maps of the LULC were more accurate than any of the four models. This was expected because combining multiple models into a hybrid model allows for the creation of a more powerful and accurate model by leveraging the diverse strengths and weaknesses of different models as they capture distinct parts of the data and

provide complementary predictions (Bagui, 2005; Chen, Dou and Yang, 2017; Polikar, 2006; Rokach, 2010). These results indicate that the Quad hybrid model could be an adaptable tool in enhancing the accuracy of modelling LULC changes for large-area mapping of wetlands, where there are limitations on resources and time coupled with challenges in gathering a large number of high-quality reference samples for calibration.

6.1.3 Land Use/Land Cover Trend Analysis from 1990 to 2020

We employed the Quad hybrid model to understand the dynamic of the LULC of the Bangweulu Wetland and the surrounding area between 1990 and 2020. The trend analysis of LULC from 1990 to 2020 using the Quad hybrid model (Table 12 and Figures 26, 27 and 28) reveals several notable patterns as elaborated below.

There is a consistent decline in forest cover, water bodies, and grassland throughout the analysed period, this kind of trend was also observed by Muche *et al.* (2023). On the contrary, there is an upward trend in cropland and built-up areas the trend which was also observed by Gxokwe, Dube and Mazvimavi (2023) and Taiwo *et al.* (2023). The decrease in the extent of forest cover and grassland could be mainly attributed to factors such as deforestation, urbanisation, and agricultural expansion, resulting in the conversion of natural land into agricultural fields and built-up infrastructure. The declining forest and grassland cover and increase in built-up and cropland areas in the BWS and the surrounding areas will have several impacts on both the wetland ecosystem itself and the surrounding environment. These impacts include (1) water quality deteriorations as observed by several other studies (Camara, Jamil and Abdullah, 2019; Song *et al.*, 2020; Tahiru, Doke and Baatuuwie, 2020; Nkwanda *et al.*, 2021; Pandey, Kumari and Al Nawajish, 2023), (2) loss of biodiversity and habitat (Fenta *et al.*, 2020; Perennou *et al.*, 2020; Meng *et al.*, 2023), (3) increased wetland sedimentation (Hernández-Romero *et al.*, 2022), and (4) alterations in the microclimate such as variation of surface temperature (AlDousari *et al.*, 2023). In addition, the observed decrease in the surface water bodies, forest cover and grassland cover will accelerate the increasing patterns of drought severity in the BWS and the surrounding areas (Kafy *et al.*, 2023).

6.2 Rapid evaluation of the variability of water quality.

6.2.1 Modified Normalised Difference Water Index (MNDWI)

The observed MNDWI values in the study area ranging from 0 to 1 (Figure 29A) were within the ideal range (Xu, 2006; Adhikari, 2019). However, the observation might be different depending on the nature of the study area, which emphasises the need to interpret the MNDWI value thresholds within the local context as the MNDWI might be influenced by various factors

such as environmental conditions, vegetation cover, and land surface properties which should be considered to avoid misinterpretation or overlooking site-specific distinctions (Du *et al.*, 2016; Szabó, Gácsi and Balázs, 2016; Laonamsai *et al.*, 2023).

6.2.2 In-situ Water quality parameters

The water quality parameters measured both on-site and, in the laboratory, (except for feacal coliform) for the Bangweulu Wetland, as shown in Table 7, were found to be lower than the established local ambient water quality standards as well as internationally accepted standards for drinking water. This supports the earlier conclusion made by Bos and Ticheler (1996) that the Bangweulu lakes are among the most dilute water bodies in Africa. The most dilute water bodies in this case refer to aquatic environments that have very low concentrations of dissolved salts (sodium and chloride) and minerals (TDS, hardness, EC, pH, and iron) (Kolding, 2011; Warren *et al.*, 2019; Huang *et al.*, 2020; Turunen *et al.*, 2020). The lower concentration levels of the water quality parameters in water may have positive implications for the local ecosystem and various human activities, including agriculture, aquaculture, recreation, industry, electrical power generation, and human consumption (United Nations Environment Programme, 2008). Specifically, the measured levels of sodium, chloride, EC, and TDS were found to be lower than both worldwide and local ambient water quality standards (Table 7). This implies that the water is fresh, clean, relatively pure, and contains a low concentration of dissolved ions and contaminants. It also indicates a lower mineral content, which is desirable for aquatic ecosystems and various uses such as drinking water supply and ecological preservation (Bhateria and Jain, 2016; Sallam and Elsayed, 2018; Aljoborey and Abdulhay, 2019).

These findings could benefit the local population who rely on the wetland waters for their livelihood (Aburto *et al.*, 2013). Besides, the lower levels of sodium and chloride could have positive effects on the ecosystem and aquatic life by reducing harm to aquatic plants, animals, and soil quality near water sources (Nielsen *et al.*, 2003). However, the water may lack other essential minerals and elements necessary for human health, such as calcium, magnesium, and fluoride (Herbert *et al.*, 2015). This is also supported by the observed lab-measured (Table 7) low calcium hardness and total hardness concentration levels, indicating a low presence of calcium and magnesium in the Bangweulu Wetland lakes (Diggs and Parker, 2009). Additionally, even though the low TDS level indicates good water quality, the lack of many minerals in the water may cause it to have a flat taste (Lestari *et al.*, 2023). Very low turbidity levels were observed as well, suggesting fewer suspended solids and purer water with a lower likelihood of bacterial infection (Allen, Copes and Hrudey, 2008). Moreover, the water's pH of

6.68 and the lack of detectable concentrations of nutrients (nitrates, phosphates, and sulphates) and heavy metals (iron, lead, and manganese) point to a decline in eutrophication and algae development. According to Singh, Kaur and Katnoria (2017) and (Wu *et al.* (2022), these results show good water quality, which is essential for maintaining the biodiversity of wetland ecosystems and ensuring the availability of supplies of safe drinking water. To make sure that these water quality parameters' concentration levels remain within safe limits and do not increase over time as a result of changing environmental or human activity, these parameters must be continuously monitored.

In some sampling points, elevated levels of feacal coliforms were observed, exceeding both local and international recommended limits (Appendix II). This indicates a significant contamination problem that requires immediate action to protect public health, preserve the environment, and secure access to safe drinking water for over fifty thousand people living in the Bangweulu Wetland (Holcomb *et al.*, 2020; Niyoyitungiye, Giri and Ndayisenga, 2020). The presence of faecal coliforms in the lakes of the Bangweulu Wetland could be linked to the numerous fishing camps dispersed throughout the wetland area. However, proper identification of the source of faecal contamination in the Bangweulu Wetland is needed to facilitate more effective interventions and halt pathogen transmission (Holcomb *et al.*, 2020).

6.2.3 Normalised Difference Turbidity Index (NDTI)

The observed NDTI values ranging from -0.363 to 0.250 (Figure 29B) in the Bangweulu water bodies imply that the water quality in these water bodies varies depending on the specific location and time. The variations could have been due to various factors such as vegetation, sedimentation, flocculation, and river inflow (Bid & Siddique, 2019; Meena et al., 2021). This could be true because high values of NDTI were observed at the edges of the western, southern, and northern sides of the water bodies, where the water bodies are shallow and the majority of the river inflow as well as the fish camps and agricultural fields are located, which could have led to high sedimentation and flocculation (Bid and Siddique, 2019; Meena *et al.*, 2021).

6.2.4 Normalized Difference Salinity Index (NDSI)

The NDSI map is an invaluable resource for tracking salinity variations over time and comprehending salinity distribution in water bodies, which is essential for several agricultural and environmental applications (Gerardo & de Lima, 2022). The NDSI values below zero (Figure 29C) suggest that the ratio of near-infrared light to red light is relatively high, which could arise from different environmental factors. In the case of aquatic bodies like the Bangweulu Wetland lakes, the observed NDSI values below zero could signify the absence of

salty water or low salinity levels (Ould Ahmed, Inoue and Moritani, 2010; Shadrin, Stetsiuk and Anufriieva, 2022). Additionally, the presence of sediments, organic matter, or other non-saline materials in the water could also lead to negative NDSI values, indicating the dominance of non-saline components in the area (Ould Ahmed, Inoue and Moritani, 2010; Shadrin, Stetsiuk and Anufriieva, 2022).

The small portions of NDSI values of 0.7 in Figure 29C observed on the edges of the lake might not necessarily indicate higher salinity levels but rather the misclassification of sand beaches as water bodies by the MNDWI algorithm or mixed pixels of water and sand due to shallow waters at the edges of the lake. The edges of lakes with sand beaches can have similar spectral reflectance properties to those of water bodies, which can lead to misclassification by the MNDWI algorithm. This is because the MNDWI uses the difference between the green and middle infrared (MIR) bands to distinguish water from land. However, sand beaches can have similar spectral signatures to water, making it difficult for the MNDWI to accurately identify the shoreline (Zhang *et al.*, 2011; Abdelhady *et al.*, 2022; Zhou *et al.*, 2023). The observed unusual phenomena of higher values of NDSI in the middle areas of Bangweulu Lake, where the lake's depth is approximately 10 meters deep (deepest point), indicating higher salinity levels and lower values of NDSI in the northern to southeastern side of the lake where the lake is shallow (approximately 4 meters). This phenomenon could be attributed to a combination of factors. One possible explanation is the limited freshwater inputs in the central part of the lake, which is far from freshwater sources (such as rivers or streams) but more fresh water inputs in the Northen to southeastern sides of the lake. This could result in less dilution of salts in the middle and more dilution in the northern and southeastern side of the lake, leading to higher salinity and lower sanity levels respectively (Hamdani *et al.*, 2018; Suchan and Azam, 2021). This could be true because the Bangewulu lake is fed by several rivers draining from the northwest slopes of the Muchinga Escarpment and the southeast slopes of the rift valley highlands, including the Chambeshi River, which flows into the lake from the north (Mccann, 2017). However, these freshwater inputs are far from the middle of the lake where the higher salinity readings were observed but very close to where lower levels of NDSI values were observed. Another factor contributing to the higher salinity in the lake's central part could be due to the presence of salts from the lake bed. If the lake bed contains mineral deposits or evaporite minerals, they can dissolve into the water column, increasing salinity. This is more likely to occur in the deeper, central parts of the lake (Hamdani *et al.*, 2018). This is supported by the (British Geological Survey, 2001), which noted that the lake bed of Bangweulu Lake in Zambia contains mineral deposits, including evaporite minerals from the Katanga System. additionally, it could also be due to anthropogenic sources as the highlands in the middle and

in between the lakes are settlements, where the major activities for people's livelihood are farming and fishing as indicated by (Chundu *et al.*, 2024).

6.2.5 EC map from the relationship with NDSI

The observed R^2 value of 0.734 (Figure 30D) for the relationship between NDSI and EC was lower than the value of R^2 of 0.87 reported by Al-Jabri *et al.* (2023) for their study conducted in Sultanate, Oman, in the summer of 2017, but in line with the R^2 of 0.766 found by the same study in the summer of 2019. The observed linear relationship between water EC and the NDSI provides valuable insights into the association between salinity and EC, with implications for various fields, including hydrology, environmental monitoring, agriculture, water resource management, and geochemistry (Corwin and Yemoto, 2017). Approximately 26.6% of the variation in water EC could not be explained by the model, either as a result of model constraints, temperature variation during data collection, dissolved ion concentration, pressure, or variations in the water sample location (Pal *et al.*, 2015; Ding *et al.*, 2018).

6.2.6 Sodium and Chloride maps from the relationship with NDSI

The study found a positive linear relationship (R^2) of 0.731 and 0.709 between sodium-NDSI and chlorides-NDSI, respectively (Figure 30: A and C), indicating that NDSI could accurately reflect the salinity of the wetland's open water bodies. This might have some important implications for assessing the wetland's water quality and environmental conditions. Factors such as rainfall patterns, groundwater flow, water pollution, temperature, evaporation, and human activities might have affected the chlorides-NDSI and sodium-NDSI relationships (Hintz and Relyea, 2019; Dey and Vijay, 2021; Ladwig, Rock and Dugan, 2023).

6.2.7 TDS maps from the relationship with EC

In water quality studies, the relationship between TDS and EC is frequently expressed using the linear model TDS = k*EC. The model's good explanatory ability in predicting TDS based on EC is indicated by the observed R^2 value of 0.907 (Figure 31A) when predicted TDS was compared with on-site measured TDS. The results are in line with other researchers (Pal *et al.*, 2015; Rusydi, 2018). This demonstrates the usefulness of converting EC maps into TDS maps using the equation TDS = k*EC.

6.2.8 Turbidity map from the relationship with NDTI

The relationship between lab-measured water turbidity and the NDTI was observed with a linear model (R^2 = 0.615) as indicated in Figure 30B. The study conducted in Panchet Hill Dam, India by Bid and Siddique (2019) observed a better correlation of R^2 of 0.866 with a linear model for the pre-monsoon period. The differences in the findings could be due to differences in the study

area as the composition of water quality parameters is site-specific due to various factors such as local geology, climate, land use, and human activity within a region (Hamid, Bhat and Jehangir, 2020). The difference in results could also be attributed to the temporal lag between in situ data collection and satellite overpass (Jordan *et al.*, 2023). The model could not explain about 38.5% turbidity variations in the water bodies. The complexity of the relationship between turbidity and NDTI could be possibly due to the influence of factors such as environmental conditions, water composition, and depth, land use, and human activities (Bid and Siddique, 2019; Rodríguez-López *et al.*, 2021). Understanding the dynamics of these factors will improve the interpretation and application of absorbance values derived from remote sensing indices in assessing water quality.

6.2.9 Water Quality Index (WQI)

Comparing water quality measurements obtained through on-site/lab methods with those derived from remote sensing reveals that the water quality parameter values derived from the in-situ/lab consistently displayed lower values compared to those derived from remote sensing (Table 14). Additionally, the WQI calculated from on-site/laboratory-measured parameters was also lower than that which was derived from remotely sensed data (as indicated in Table 14), this phenomenon was also highlighted by Adjovu *et al.* (2023). The discrepancy could be attributed to the unique capabilities of remotely sensed data. Remote sensing often covers wider areas and provides synoptic views of water bodies, which results in average values that do not adequately reflect the local conditions (localised measurements/point measurements) recorded by the on-situ/lab observations. Furthermore, remote sensing technologies may be affected by atmospheric interference, which can result in biases in estimated water quality parameters leading to capturing localised water quality variations that may not match that of on-site/lab measurements (Gholizadeh *et al*., 2016; Adjovu *et al.*, 2023). In addition, although in situ water quality observations are generally considered more reliable and accurate than remote sensing water quality observations (Murray *et al.*, 2022; Kumar *et al.*, 2024), the accuracy and precision of the field observation data may be problematic due to errors in field sampling and laboratory procedures (Sputh *et al.*, 2008; Gholizadeh *et al*., 2016), which could also have contributed to the observed discrepancy. Despite the numerical differences in WQIs, both data sets fell within the 'Good' category (Tables 8 and 14) implying that the water of the Bangweulu Wetland lakes is suitable for drinking, irrigation, and industrial purposes according to the water quality status classification (Table 8). Furthermore, the observation suggests that both approaches captured the essential aspects of water quality and can complement each other effectively which was also observed by Arabi *et al.* (2020) in the Dutch Wadden sea. In light of these observations,

decision-makers should consider leveraging the interactions between on-site/lab and remote sensing methods to better monitor and manage water quality as each approach has its unique advantages and limitations, and their combined use could lead to more informed decisions (Gholizadeh *et al*., 2016; Arabi *et al.*, 2020).

6.2.10 Reclassification and Integration of Different Water Quality Maps

Based on the Water Quality Index (WQI) value range (Figures 34 and 35), areas with values from 0 to 25 are classified as "excellent," indicating well-protected water quality, suitable for various uses, and might not require immediate action. Whereas, values ranging from 26 to 50 are classified as "good," with minor threats but still safe for consumption and industrial use, and could indicate a level of water quality that might need management and monitoring (Brown *et al.*, 1970; Islam, Azadi and Nasiruddin, 2021; Chidiac *et al.*, 2023). Values from 51 to 75 are classified as "poor," showing occasional vulnerability to pollution, suitable for irrigation and industrial purposes, and could indicate areas where water quality pollution activities might be active. Values from 76 to 100 are classified as "very poor," frequently vulnerable to pollution, and suitable only for irrigation, and could indicate areas where water quality pollution activities might be very active, necessitating immediate action to avoid future degradation and safeguard human health and the wetland ecosystem (Brown *et al.*, 1970; Islam, Azadi and Nasiruddin, 2021; Chidiac *et al.*, 2023). In Figure 35, I observed that the “poor” class (Values from 51 to 75), was found both in certain sections of the lake shores as well as middle of the lake. The observed poor water quality in certain sections of the lake shores could be attributed to several factors. One significant contributor is the shallow water and reduced mixing which allows pollutants to accumulate in the nearshore areas rather than being dispersed throughout the lake, as noted by Bhateria and Jain (2016). Another factor is the spread of water hyacinth in the shores of the lake (Bhateria and Jain, 2016) as well as sedimentation and turbidity due to wind-driven waves and anthropogenic activities, as observed in Lake Pontchartrain, Louisiana, United States of America by Hyun (2007). In addition, localized pollution due to shoreline communities which often directly discharge waste into the lake could also be a contributing factor (Karmakar and Mavukkandy, 2013). Whereas, the observed poorer water quality in the middle section of the lake compared to the shores could be attributed to several factors which have been partly alluded to in section 6.2.4. In addition, the water circulation patterns could play a crucial role. Surface water can sink to the bottom of the lake, carrying pollutants and nutrients. As these nutrients are distributed throughout the lake, the middle sections, where water circulation is less intense, may receive higher concentrations, leading to poorer water quality compared to the shores (Wang *et al.*, 2020). Furthermore, biological processes such as

algal blooms could also impact water quality. These blooms consume oxygen, leading to hypoxia, which is more pronounced in the middle sections where water circulation is less intense (Bhateria and Jain, 2016). Besides, Sedimentation could be another contributor, as polluted sediments carried by runoff can settle at the bottom of the lake, particularly in the middle sections as observed by Bhateria and Jain, (2016). Therefore, categorising water body areas through integrated WQI classes could help water resource managers, policymakers, and researchers make informed decisions about water use and management. It could also assist in assessing water quality health risks, planning water management strategies, prioritising wetland ecosystem conservation efforts, and understanding human activity's environmental impact on water quality. Additionally, this approach could aid in identifying regions needing immediate action to improve water quality and safeguard the welfare of communities relying on wetland ecosystem services. Furthermore, it could also help prioritise interventions based on the water quality severity of the inland open water bodies.

6.3 Assessment of the Linkages Between Land-Use/Land-Cover and Water Quality

6.3.1 Influence of LULC on water quality parameters

The findings in this study through multiple regression analysis (refer to Table 17 above) highlight significant correlations between some water quality parameters (Turbidity, TDS, Fe^{2+}, and EC) and LULC with R^2 values exceeding 0.7, significance F values below 0.05, and positive coefficients. This indicates that these water quality parameters are strongly influenced by the characteristics of the LULC in the study area which was observed by other studies (Afed Ullah et al., 2018; Li et al., 2008; Nguvulu et al., 2021). The high R^2 values suggest that a large portion of the variation in Turbidity, TDS, Fe^{2+}, and EC can be explained by the LULC. Additionally, the low significance F values (below 0.05) demonstrate that the relationship between these water quality parameters and LULC is statistically significant, meaning that the observed correlation is unlikely to have occurred by chance. The positive coefficients further confirm that changes in LULC directly impact the levels of Turbidity, TDS, Fe^{2+}, and EC in the water. Therefore, this study has established that Turbidity, TDS, Fe^{2+}, and EC are the most sensitive water quality parameters to changes in LULC in this study area, as their strong and statistically significant correlations with LULC suggest that the characteristics of the LULC play a crucial role in determining the levels of these water quality parameters. These findings emphasise the critical role of considering spatial dynamics and land use characteristics when assessing and managing water quality in the study area, as alterations in LULC could have a significant impact on certain key water quality indicators which was also observed by Rather & Dar (2020) and Tahiru et al. (2020).

Further analysis revealed a specific strong correlation between turbidity levels and built-up areas, supported by a p-value below 0.05 and a positive coefficient, indicating a likely non-random correlation. This is in line with Nguvulu et al. (2021) who also observed a significant relationship between these variables with a p-value of less than 0.05. The positive coefficient observed suggests that increased development in an area corresponds to higher turbidity levels in nearby water bodies, possibly due to factors like increased runoff, erosion, and pollutant release from human activities in built-up areas. Chen & Chang (2019) also observed that urban monitoring stations located in highly developed areas had elevated turbidity levels than in less developed areas. Similarly, the correlation between Fe^{2+} levels in water and forested areas, with a significant p-value of less than 0.05, indicates a meaningful relationship not attributable to chance. The positive coefficient implies that as forested areas expand, Fe^{2+} levels in water tend to rise, likely due to forests acting as reservoirs of minerals like Fe^{2+} and facilitating its transport into water bodies through natural processes as also observed by Björnerås et al. (2017) who concluded that the widespread increase in Fe^{2+} concentrations in European and North American freshwaters were associated with expanding forestry. Additionally, Nguvulu et al. (2021) in his conclusion indicated that forest cover had a negative correlation with most water quality parameters except for Fe^{2+}, though it was not statistically significant.

The correlation between TDS, EC, and cropland, with a p-value below 0.05 and a positive coefficient, highlights a significant link between agricultural activities in the study area and the two water quality parameters (TDS and EC). The observed positive coefficient suggests that as cropland extent increases, TDS and EC levels in water also increase, potentially influenced by agricultural practices such as the application of chemical fertilizer, organic manure and pesticide as well as irrigation activities and runoff from croplands which could introduce dissolved salts, minerals, and other ions into the water (Hu *et al.*, 2019; Rey-Romero, Domínguez and Oviedo-Ocaña, 2022). However, parameters like K^+, Na^+, Cl-, and Ca^{2+} exhibited a low correlation with LULC, with R^2 values less than 0.1, implies that these specific parameters are less influenced by the LULC characteristics in the study area. This suggests that variations of these water quality parameters are likely driven by factors like hydrology, geology, soil types, slope and anthropogenic activities other than the spatial patterns of LULC (Karakuş, 2020; Li et al., 2008; Li et al., 2012). Therefore, while factors like turbidity, TDS, Fe^{2+}, and EC show a strong relationship with LULC, the limited correlation of K^+, Na^+, Cl^-, and Ca^{2+} with LULC highlights the complexity of factors influencing various water quality parameters. This finding underscores the need to consider a broader range of factors beyond LULC when

assessing and managing the quality of water resources in the study area. This includes understanding the complex interactions between LULC, hydrology, geology, and other anthropogenic (fishing activities, reed harvesting and seasonal fish camps) activities to develop effective strategies for maintaining or improving water quality in this study area (Karakuş, 2020; Li et al., 2008; Li et al., 2012).

Feacal coliform had a relatively lower R^2 of 0.258 with LULC, implying a weaker correlation between feacal coliform levels and LULC. A lower R^2 value indicates that the variability in feacal coliform levels is not well explained by the different LULC types. The significance F value of 0.063 suggests that the relationship between feacal coliform and LULC might be statistically significant because it was so close to 0.05. further analysis of the relationship between feacal coliform and specific LULC types revealed a lower p-value of 0.006 for built-up areas, with a positive coefficient, indicating that the correlation between feacal coliform and built-up areas was statistically significant. The positive coefficient implies that as the extent of built-up areas increases, feacal coliform levels also tend to increase. This is likely due to increased human activities, domestic wastewater, and sewage discharge in built-up areas. Various studies have also consistently demonstrated that increased urbanisation, built-up areas, and population density are strongly correlated with higher levels of feacal coliform bacteria in water sources (Nogueira *et al.*, 2003; Yuan *et al.*, 2019; X. Zhang *et al.*, 2020).

Generally, the findings highlight the fact that understanding the relationships between water quality parameters and LULC types is crucial for water resource management. As observed from the results, it could be deduced that conservation of forested areas in wetlands is vital for maintaining mineral balance in water sources and excellent water quality as observed by Bekiroğlu & Eker (2011) and Dudley et al. (2003), while sustainable land management practices in croplands could help mitigate negative effects (TDS and EC) on water quality which was also highlighted by Convention on Wetlands (2022). Furthermore, the interventions to reduce feacal contamination so as to safeguard public health should focus on managing the impacts of urbanisation and human settlement, rather than just broad land use changes (French *et al.*, 2021).

6.3.2 Water quality index for various sampling points

The wide range of WQI values, ranging from 32.941 to 255.543 (refer to Figure 36) across the 34 sampling points in the catchment, signifies significant spatial disparities in water quality within the study area as observed by other studies (Ahmed et al., 2023; Wang et al., 2017). The

observed variations could be due to differences in soil types, geology, hydrology, geographical and climatic characteristics as well as human factors (Chidiac *et al.*, 2023). As guided by Table 8 above, a WQI value below 50 indicates good quality water suitable for drinking, irrigation, and industrial purposes. Streams falling within this range, such as JK008 with a WQI of 32.94, imply water of high quality that could be used for various applications without extensive treatment. Conversely, streams with WQI values between 50 and 75 imply poor quality, suitable only for irrigation and industrial uses, not for direct human consumption, necessitating some level of treatment before domestic use. Streams with WQI values between 75 and 100 are deemed very poor quality, only suitable for irrigation post proper treatment. Streams exceeding a WQI of 100 are extremely poor (like Nkulumashiba Bridge with a WQI of 255.54), requiring significant treatment for any purpose due to high contamination levels, primarily attributed to elevated Fe^{2+} content (2.23mg/l) surpassing the ambient standard of 0.7mg/l. The notable variations in WQI across the catchment (Figure 36) suggest a diverse water quality scenario influenced by LULC characteristics as well as natural factors such as geology, hydrology, climate (Ahmed et al., 2023; Wang et al., 2017). These findings underscore the significance of comprehensive water quality monitoring and assessment across multiple locations within a catchment to pinpoint areas of concern and guide targeted management strategies. These insights are crucial for decision-makers, environmental agencies, and communities, emphasizing the need for a holistic approach that considers both natural processes and human activities in addressing water quality challenges. Streams with very poor or high WQI levels like Nkulumashiba, Chisangwa 1, JK002, and JK003, among others (refer to Figure 36) pose health risks to humans, aquatic life, agriculture, and ecosystems, necessitating focused interventions like pollution control measures and improved land management practices. Furthermore, efforts should be directed towards monitoring and managing these streams, identifying pollution sources (especially Fe^{2+}), and implementing corrective actions. In addition, promoting sustainable land use practices to safeguard water quality and collaborating with stakeholders like local communities and government agencies could be vital for effective water quality management.

6.3.3 Influence of LULC on WQI

The WQI is influenced by a combination of natural and anthropogenic factors (Akhtar *et al.*, 2021). In this study, I observed a notable correlation between LULC and the overall WQI through multiple regression analysis (Table 17) with an R^2 value of 0.649 and a positive coefficient, suggesting that LULC changes explained 64.9% of the WQI variation. The observed positive coefficient indicates that the WQI changes with changes in LULC. This

observation agrees with Tahiru et al. (2020) who also noticed a positive significant relationship between LULC types and water quality. Additionally, other studies have also observed that most water contamination occurs due to increased anthropogenic activities (Sharma and Bhattacharya, 2017; Pandey, Kumari and Al Nawajish, 2023). The 35.1% which could not be explained by the LULC factor could be attributed to natural factors, such as climate change, natural disasters, geological factors, soil composition, and hyporheic exchange, which play a significant role in shaping water quality and can impact the presence of minerals, dissolved substances, and overall water composition (Akhtar *et al.*, 2021). Further analysis showed that the WQI was more correlated with forest cover, supported by a p-value of below 0.05 (Table 17), demonstrating a significant statistical relationship between WQI and forest cover in the study area. A similar observation was made by Piyathilake et al. (2022), who also noted that the areas that were mainly forested had WQI values falling within the good and excellent status, and that the water quality deteriorated from the forested areas to the unforested terrain areas.

6.4 Policy Recommendations and Implication

There are various policy implications for the observed changes in LULC. A considerable environmental protection strategies are necessary to minimise deforestation and land degradation, given the fall in the forests and grassland covers. Policies for sustainable urban development are also required to deal with the significant rise in built-up areas which was also suggested by Dey *et al.* (2021). The necessity for agricultural and food security measures, particularly the promotion of sustainable farming methods, is highlighted by the substantial growth of cropland.

The study underscores the significance of engaging local communities in sustainable management practices. Initiatives such as sustainable fisheries, eco-tourism, and agroforestry could offer alternative livelihoods, thereby reducing the strain on Bangweulu wetland resources. Moreover, implementing environmental education programs is crucial for raising awareness about the effects of LULC changes on water quality and ecosystem health. This awareness could foster community-driven conservation efforts, particularly in forest restoration, which could be of help in counteracting the observed deforestation trends. By doing so, these strategies could enhance both land productivity and climate resilience while mitigating climate change impacts. Additionally, these strategies could also help maintain essential ecosystem services such as carbon sequestration and biodiversity conservation as observed by Phethi and Gumbo (2019).

The study also reveals a reasonable correlation between LULC changes and WQI (R^2 = 0.649) which underscores the need for integrated policies that link land use planning with water resource management. Furthermore, developing adaptive management frameworks that incorporate advanced technologies like machine learning and remote sensing could improve decision-making processes for wetland conservation.

In view of the above, the assessment of the water-body changes is critical for determining the natural and human-induced impact on the water bodies. The continuous assessment of the changes in the water body could contribute knowledge to various fields, including studies on water resource management and disaster risk reduction, by providing critical information about the health and state of freshwater ecosystems (Sarp and Ozcelik, 2017; Wang *et al.*, 2023). In addition, the long-term water-body change assessment and monitoring will be essential for evaluating the success of interventions, guiding conservation efforts, mitigating negative impacts on the wetland ecosystem, and determining whether the reduction in water in the BWS is a constant trend or a short-term phenomenon.

6.5 Study Limitation and Uncertainties

The study had several limitations and uncertainties that should be considered. Firstly, it was difficult to distinguish certain wetland's different LULC features using 30-m Landsat images, which could have led to classification errors (Commey, Magome and Ishidaira, 2023). Furthermore, it could be possible that some of the models' hyper-parameters used might not have been fully optimised, which could have affected how accurately the LULC maps were classified (Ebrahimi-Khusfi, Nafarzadegan and Dargahian, 2021). Also, the study used Landsat 5 images of 1990, 2000, and 2010, which had some sparkles in the 1990 and 2000 images, this could have led to poor performance in classifying certain LULC classes (Peng *et al.*, 2024). In addition, this study used limited training sample size in the LULC classification that might not have been adequately represented the full variability of the dataset. This could have restricted the ability of certain ML models such as the RF to generalise well to unseen instances thereby creating an unstable decision tree with reduced accuracy (Polikar, 2006; Fernández-Delgado *et al.*, 2014). Furthermore, the study excluded some models, such as the ANN model because it could not run in Orfeo Toolbox in QGIS 3.28.5, which might have limited the range of models considered for further analysis.

Secondly, this research used statistical equations to create water quality parameters and WQI maps from water quality indices and empirical equations. These equations are based on the

specific conditions of the lakes under consideration and are therefore not generic. Secondly, relying on average November 2022 imagery for analysis might not have effectively captured the full spectrum of water quality fluctuations that occurred throughout the different days of the data collection. Additionally, assuming that water quality parameter concentration levels did not change significantly between 2022 and 2023 might not be entirely accurate, as water quality can be influenced by various factors such as seasonal changes, human activities, and climate variability (Lintern *et al.*, 2018; Anh *et al.*, 2023). This might have restricted the predictive power of our models, potentially leading to incorrect conclusions.

Lastly, some sampling points (Table 16) like MC 11HM, MC 8 HM, and MC 9HM had missing data on pH, TDS, and Fe^{2+}. R2-28, R2-23, and R2-19 had missing data on Fe^{2+}. To fill the gaps in the missing data, mean imputation approach was used, where average values from nearby points were used as estimates (Zhang & Thorburn, 2022). However, this approach assumed that the nearby streams accurately represented the missing data points, which might not have been the case due to factors like watershed size and land use patterns, potentially affecting parameter concentrations. This could have compromised the statistical analysis, leading to erroneous or bias conclusions (Zhang & Thorburn, 2022).

CHAPTER SEVEN: CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

The specific conclusions of this study based on the four research questions (section 1.4) are highlighted below.

The SVM, NB, DT, and KNN were identified as the best-performing ML models for the LULC classification in the Bangweulu Wetland and the surrounding areas. The Quad hybrid model was developed, which achieved a higher level of accuracy than individual ML models (SVM, NB, DT, and KNN). The developed Quad hybrid model was used to study the past and the present (1990 to 2020) LULC changes in the Bangweulu Wetland and the surrounding area. The study has shown that there has been a consistent increase in cropland and built-up areas from 1990 to 2020 at the expense of the forest cover and grassland. The water body also experienced a gradual reduction between 1990 and 2020, although at a minimal rate. Our observation highlights significant transformations in the LULC composition over the studied time frame. This could jeopardise the ecological integrity of the BWS, thereby compromising the provision of social, economic, and environmental benefits to both human and natural systems.

Both on-site/lab and remote sensing methods indicated that the water quality in Bangweulu Wetland falls below local and international standards, with a calculated Water Quality Index (WQI) falling within the 'Good' category. This suggests that the water is fresh, clean, and suitable for various purposes including ecological preservation, agriculture, aquaculture, recreation, industry, and human consumption. Integrating multiple WQI maps into a single map allows stakeholders to prioritise interventions based on the severity of water quality issues in different locations. The findings offer insights into the general status of water quality in open water bodies, identifying hot spots and potential sources of water pollution with implications for future management practices. The study provides a promising approach to overcoming challenges associated with monitoring water quality parameters in large water bodies, aiding in the conservation of aquatic ecosystems and the supply of clean drinking water.

The study highlights the significant influence of LULC distribution on water quality, emphasising the importance of considering LULC in water quality management. Turbidity, TDS, Fe+2, and EC were identified as the most sensitive water quality parameters to changes in LULC, suggesting that LULC characteristics play a crucial role in determining these parameters in the study area. The study observed wide range of WQI values across sampling points within the catchment,

indicating spatial disparities in water quality. The study established that WQI changes in response to LULC changes, but also noted that natural factors significantly influence water quality. This underscores the need to consider a broader range of factors beyond LULC when assessing and managing water quality. Recognising the relationships between LULC and water quality could improve decision-making on land use planning and water quality management, enabling policymakers, environmental planners and managers to predict river water quality and reduce the need for periodic sampling processes. This could be particularly useful for shaping sustainable LULC policies, identifying hotspots of potential water quality degradation, and pinpointing areas for targeted interventions and restoration efforts.

7.2 Recommendations

Drawing from the findings of this study, the following recommendations are made to water resource managers in both public and non-governmental organisations, as well as researchers.

- To validate the performance consistency of the Quad hybrid model, which is being used for the first time, it is proposed that the University of Zambia through IWRM centre should evaluate the model's robustness by assessing its performance across different time periods, geographic locations, and data types to determine how consistently it performs under varying conditions.

- The study highlights the need by the government and African Parks for effective land management strategies and sustainable planning to mitigate the potential impacts on natural ecosystems and ensure a balance between human activities and environmental conservation of the BWS.

- The water resource managers and researchers should consider the potential use of remote sensing techniques for rapid monitoring of water quality parameters variabilities in inland open water bodies at a large scale. As the study demonstrates that existing empirical models and spatial data could aid in identifying regions needing immediate action to improve water quality.

- Water resource managers and researchers should take advantage of the complementary strengths of on-site/laboratory and remote sensing methods in assessing water quality variability to enhance their monitoring and management.

REFERENCES

Abdelhady, H.U. *et al.* (2022) 'A Simple, Fully Automated Shoreline Detection Algorithm for High-Resolution Multi-Spectral Imagery', *Remote Sensing*, 14(3), p. 557. Available at: https://doi.org/10.3390/rs14030557.

Abdi, A.M. (2020) 'Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data', *GIScience and Remote Sensing*, 57(1), pp. 1–20. Available at: https://doi.org/10.1080/15481603.2019.1650447.

Abidi, S.M.R. *et al.* (2020) 'Educational sustainability through big data assimilation to quantify academic procrastination using ensemble classifiers', *Sustainability (Switzerland)*, 12(15), pp. 1–23. Available at: https://doi.org/10.3390/su12156074.

Aburto, N.J. *et al.* (2013) 'Effect of lower sodium intake on health: Systematic review and meta-analyses', *BMJ (Online)*, 346(7903), pp. 1–20. Available at: https://doi.org/10.1136/bmj.f1326.

Acharya, T.D. *et al.* (2017) 'Combining Water Indices for Water and Background Threshold in Landsat Image', in. MDPI AG, p. 143. Available at: https://doi.org/10.3390/ecsa-4-04902.

Adhikari, S. (2019) 'A Comparison of Water Indices and Binary Thresholding Techniques for Water Surface Delineation for St. Croix Watershed Area', *Yearbook of the Association of Pacific Coast Geographers*, 81(1), pp. 182–204. Available at: https://doi.org/10.1353/pcg.2019.0002.

Adjovu, G.E. *et al.* (2023) 'Overview of the Application of Remote Sensing in Effective Monitoring of Water Quality Parameters', *Remote Sensing*. Multidisciplinary Digital Publishing Institute, p. 1938. Available at: https://doi.org/10.3390/rs15071938.

Adugna, T., Xu, W. and Fan, J. (2022) 'Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images', *Remote Sensing*, 14(3), pp. 1–22. Available at: https://doi.org/10.3390/rs14030574.

Afed Ullah, K., Jiang, J. and Wang, P. (2018) 'Land use impacts on surface water quality by statistical approaches', *Global Journal of Environmental Science and Management*, pp. 231–250. Available at: https://doi.org/10.22034/gjesm.2018.04.02.010.

Ahmad, W. *et al.* (2021) 'Impact of land use/land cover changes on water quality and human health in district Peshawar Pakistan', *Scientific Reports*, 11(1), pp. 1–14. Available at: https://doi.org/10.1038/s41598-021-96075-3.

Ahmed, S.A., Dey, S. and Sarma, K.K. (2011) 'Image texture classification using Artificial Neural Network (ANN)', in *Proceedings - 2011 2nd National Conference on Emerging Trends and Applications in Computer Science, NCETACS-2011*. IEEE, pp. 56–59. Available at: https://doi.org/10.1109/NCETACS.2011.5751383.

Ahmed, W. *et al.* (2023) 'Tigris River water surface quality monitoring using remote sensing data and GIS techniques', *Egyptian Journal of Remote Sensing and Space Science*, 26(3), pp. 816–825. Available at: https://doi.org/10.1016/j.ejrs.2023.09.001.

Akhtar, N. *et al.* (2021) 'Various natural and anthropogenic factors responsible for water quality degradation: A review', *Water (Switzerland)*. Multidisciplinary Digital Publishing Institute, p. 2660. Available at: https://doi.org/10.3390/w13192660.

Al-Jabri, K. *et al.* (2023) 'Remote Sensing Analysis for Vegetation Assessment of a Large-Scale Constructed Wetland Treating Produced Water Polluted with Oil Hydrocarbons', *Remote Sensing*, 15(24). Available at: https://doi.org/10.3390/rs15245632.

AlDousari, A.E. *et al.* (2023) 'Summertime Microscale Assessment and Prediction of Urban Thermal Comfort Zone Using Remote-Sensing Techniques for Kuwait', *Earth Systems and Environment*, 7(2), pp. 435–456. Available at: https://doi.org/10.1007/s41748-023-00340-6.

Aljoborey, A.D.A. and Abdulhay, H.S. (2019) 'Estimating total dissolved solids and total suspended solids in mosul dam lake in situ and using remote sensing technique', *Periodicals of Engineering and Natural Sciences*, 7(4), pp. 1755–1767. Available at: https://doi.org/10.21533/pen.v7i4.832.

Allbed, A., Kumar, L. and Aldakheel, Y.Y. (2014) 'Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region', *Geoderma*, 230–231, pp. 1–8. Available at: https://doi.org/10.1016/j.geoderma.2014.03.025.

Allen, M.J., Copes, R. and Hrudey, S. (2008) 'Turbidity and microbial risk in drinking water Environmental analysis View project No project View project', (March). Available at: https://www.researchgate.net/publication/228605563.

Alikhani, S., Nummi, P., & Ojala, A. (2021). Urban wetlands: A review on ecological and cultural values. Water (Switzerland), 13(22), 1–47. https://doi.org/10.3390/w13223301

Alshari, E.A. and Gawali, B.W. (2022) 'Modeling Land Use Change in Sana'a City of Yemen with MOLUSCE', *Journal of Sensors*, 2022. Available at: https://doi.org/10.1155/2022/7419031.

Anguita, D. *et al.* (2010) 'Model selection for support vector machines: Advantages and disadvantages of the Machine Learning Theory', *Proceedings of the International Joint Conference on Neural Networks* [Preprint], (August). Available at: https://doi.org/10.1109/IJCNN.2010.5596450.

Anh, N.T. *et al.* (2023) 'Influences of key factors on river water quality in urban and rural areas: A review', *Case Studies in Chemical and Environmental Engineering*, 8, p. 100424. Available at: https://doi.org/10.1016/j.cscee.2023.100424.

Antunes, E. *et al.* (2021) 'Application of biochar for emerging contaminant mitigation', in *Advances in Chemical Pollution, Environmental Management and Protection*. Elsevier, pp. 65–91. Available at: https://doi.org/10.1016/bs.apmp.2021.08.003.

Arabi, B. *et al.* (2020) 'Integration of in-situ and multi-sensor satellite observations for long-term water quality monitoring in coastal areas', *Remote Sensing of Environment*, 239(November 2019), p. 111632. Available at: https://doi.org/10.1016/j.rse.2020.111632.

De Araujo Barbosa, C.C., Atkinson, P.M. and Dearing, J.A. (2015) 'Remote sensing of ecosystem services: A systematic review', *Ecological Indicators*, 52, pp. 430–443. Available at: https://doi.org/10.1016/j.ecolind.2015.01.007.

Atekwana, Eliot A. *et al.* (2004) 'The relationship of total dissolved solids measurements to bulk electrical conductivity in an aquifer contaminated with hydrocarbon', *Journal of Applied Geophysics*, 56(4), pp. 281–294. Available at: https://doi.org/10.1016/J.JAPPGEO.2004.08.003.

de Baar, H.J.W., van Heuven, S.M.A.C. and Middag, R. (2018) 'Ocean salinity, major elements, and thermohaline circulation', in *Encyclopedia of Earth Sciences Series*. Springer Netherlands, pp. 1042–1048. Available at: https://doi.org/10.1007/978-3-319-39193-9_120-1.

Bagui, S.C. (2005) *Combining Pattern Classifiers: Methods and Algorithms*, *Technometrics*. Available at: https://doi.org/10.1198/tech.2005.s320.

Bahrawi, J.A. and Elhag, M. (2019) 'Consideration of seasonal variations of water radiometric indices for the estimation of soil moisture content in arid environment in Saudi Arabia', *Applied Ecology and Environmental Research*, 17(1), pp. 285–303. Available at: https://doi.org/10.15666/aeer/1701_285303.

Ballesteros-Pérez, P., González-Cruz, M.C. and Mora-Melià, D. (2018) 'EXPLAINING THE BAYES' THEOREM GRAPHICALLY', in *INTED2018 Proceedings*. IATED, pp. 1608–1614. Available at: https://doi.org/10.21125/inted.2018.0028.

Banda, A.M. *et al.* (2023) 'Assessment of land use change in the wetland of Barotse Floodplain, Zambezi River Sub-Basin, Zambia', *Natural Hazards*, 115(2), pp. 1193–1211. Available at: https://doi.org/10.1007/s11069-022-05589-0.

Bangweulu | African Parks (2021). Available at: https://www.africanparks.org/the-parks/bangweulu (Accessed: 29 April 2022).

Basheer, S. *et al.* (2022) 'Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques', *Remote Sensing*, 14(19), pp. 1–18. Available at: https://doi.org/10.3390/rs14194978.

Bayas, S. *et al.* (2022) 'Land Use Land Cover Classification Using Different ML Algorithms on Sentinel-2 Imagery', in *Lecture Notes in Electrical Engineering*. Springer Science and Business Media Deutschland GmbH, pp. 761–777. Available at: https://doi.org/10.1007/978-981-19-0840-8_59.

Bekiroğlu, S. and Eker, O. (2011) 'The importance of forests in a sustainable supply of drinking water: Istanbul example', *African Journal of Agricultural Research*, 6(7), pp. 1794–1801. Available at: https://www.internationalscholarsjournals.com/articles/the-importance-of-forests-in-a-sustainable-supply-of--drinking-water--istanbul-example.pdf.

Belgiu, M. and Drăgu, L. (2016) 'Random forest in remote sensing: A review of applications and future directions', *ISPRS Journal of Photogrammetry and Remote Sensing*. Elsevier, pp. 24–31. Available at: https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Belward, A.S. and Skøien, J.O. (2015) 'Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites', *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, pp. 115–128. Available at: https://doi.org/10.1016/j.isprsjprs.2014.03.009.

Bhateria, R. and Jain, D. (2016) 'Water quality assessment of lake water: a review', *Sustainable Water Resources Management*, 2(2), pp. 161–173. Available at: https://doi.org/10.1007/s40899-015-0014-7.

Bid, S. and Siddique, G. (2019) 'Identification of seasonal variation of water turbidity using NDTI method in Panchet Hill Dam, India', *Modeling Earth Systems and Environment*. Springer International Publishing, pp. 1179–1200. Available at: https://doi.org/10.1007/s40808-019-00609-8.

Björnerås, C. *et al.* (2017) 'Widespread Increases in Iron Concentration in European and North American Freshwaters', *Global Biogeochemical Cycles*, 31(10), pp. 1488–1500.

Available at: https://doi.org/10.1002/2017GB005749.

Boateng, E.Y., Otoo, J. and Abaye, D.A. (2020) 'Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review', *Journal of Data Analysis and Information Processing*, 08(04), pp. 341–357. Available at: https://doi.org/10.4236/jdaip.2020.84020.

Bos, A. and Ticheler, H. (1996) 'A Limnological Update of the Bangweulu Fishery, Zambia', p. 28. Available at: https://www.researchgate.net/publication/285586646_A_Limnological_Update_of_the_Bangweulu_Fishery_Zambia (Accessed: 6 November 2023).

Breiman, L. (2001a) 'Random forests', *Machine Learning*, 45(1), pp. 5–32. Available at: https://doi.org/10.1023/A:1010933404324/METRICS.

Breiman, L. (2001b) 'Statistical modeling: The two cultures', *Statistical Science*, 16(3), pp. 199–215. Available at: https://doi.org/10.1214/ss/1009213726.

British Geological Survey (2001) *Groundwater Quality: Zambia Background*. Available at: https://nora.nerc.ac.uk/id/eprint/516326/1/Zambia.pdf (Accessed: 10 June 2024).

Brown, R.M. *et al.* (1970) 'A-Water-Quality-Index-Do-we-dare-BROWN-R-M-1970', *Water Sewage Works*, pp. 339–343. Available at: https://idoc.pub/documents/a-water-quality-index-do-we-dare-brown-rm-1970-6ng25k6e91lv.

Camara, M., Jamil, N.R. and Abdullah, A.F. Bin (2019) 'Impact of land uses on water quality in Malaysia: a review', *Ecological Processes*. Springer Verlag, pp. 1–10. Available at: https://doi.org/10.1186/s13717-019-0164-x.

Camargo, F.F. *et al.* (2019) 'A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images', *Remote Sensing*, 11(13). Available at: https://doi.org/10.3390/rs11131600.

Canadian Council of Ministers of the Environment. (2001) 'Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, Technical Report', *Canadian Council of Ministers of the Environment 2001. Publication No. 1299; ISBN 1-896997-34-1*, pp. 1–13.

Caruana, R. and Niculescu-Mizil, A. (2006) 'An empirical comparison of supervised learning algorithms', *ACM International Conference Proceeding Series*, 148(June), pp. 161–168.

Available at: https://doi.org/10.1145/1143844.1143865.

Casal, G. (2022) ‘Assessment of Sentinel-2 to monitor highly dynamic small water bodies: The case of Louro lagoon (Galicia, NW Spain)’, *Oceanologia*, 64(1), pp. 88–102. Available at: https://doi.org/10.1016/j.oceano.2021.09.004.

Chao, X. *et al.* (2022) ‘An efficiency curve for evaluating imbalanced classifiers considering intrinsic data characteristics: Experimental analysis’, *Information Sciences*, 608, pp. 1131–1156. Available at: https://doi.org/10.1016/j.ins.2022.06.045.

Charif, O. *et al.* (2012) ‘Cellular automata model based on machine learning methods for simulating land use change’, *Proceedings - Winter Simulation Conference*, pp. 1846–1857. Available at: https://doi.org/10.1109/WSC.2012.6465098.

Chawla, I., Karthikeyan, L. and Mishra, A.K. (2020) ‘A review of remote sensing applications for water security: Quantity, quality, and extremes’, *Journal of Hydrology*, 585(March), p. 124826. Available at: https://doi.org/10.1016/j.jhydrol.2020.124826.

Chen, J. and Chang, H. (2019) ‘Dynamics of wet-season turbidity in relation to precipitation, discharge, and land cover in three urbanizing watersheds, Oregon’, *River Research and Applications*, 35(7), pp. 892–904. Available at: https://doi.org/10.1002/rra.3487.

Chen, S. *et al.* (2020) ‘A novel selective naïve Bayes algorithm’, *Knowledge-Based Systems*, 192, p. 105361. Available at: https://doi.org/10.1016/j.knosys.2019.105361.

Chen, Y.-P.P. *et al.* (2010) ‘Bioinformatics’, *Comprehensive Natural Products II*, pp. 569–593. Available at: https://doi.org/10.1016/B978-008045382-8.00729-2.

Chen, Y., Dou, P. and Yang, X. (2017) ‘Improving land use/cover classification with a multiple classifier system using AdaBoost integration technique’, *Remote Sensing*, 9(10). Available at: https://doi.org/10.3390/rs9101055.

Cheng, C. *et al.* (2022) ‘What is the relationship between land use and surface water quality? A review and prospects from remote sensing perspective’, *Environmental Science and Pollution Research*, 29(38), pp. 56887–56907. Available at: https://doi.org/10.1007/s11356-022-21348-x.

Chicco, D., Warrens, M.J. and Jurman, G. (2021) ‘The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation’, *PeerJ Computer Science*, 7, pp. 1–24. Available at: https://doi.org/10.7717/PEERJ-CS.623.

Chidiac, S. *et al.* (2023) ‘A comprehensive review of water quality indices (WQIs): history, models, attempts and perspectives’, *Reviews in Environmental Science and Biotechnology*. Nature Publishing Group, pp. 349–395. Available at: https://doi.org/10.1007/s11157-023-09650-7.

Chittaranjan, A. (2019) ‘The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives’, pp. 210–215. Available at: https://doi.org/10.4103/IJPSYM.IJPSYM.

Chollet, F. (2017) ‘Machine learning’, *Machine Learning*, 45(13), pp. 40–48. Available at: https://doi.org/10.1007/978-1-4302-5990-9_1.

Chundu, M.L. *et al.* (2024) ‘Modeling land use / land cover changes using quad hybrid machine learning model in Bangweulu wetland and surrounding areas , Zambia’, *Environmental Challenges*, 14(February), p. 100866. Available at: https://doi.org/10.1016/j.envc.2024.100866.

Coesel, P.F.M. and van Geest, A. (2014) ‘New or otherwise interesting desmid taxa from the Bangweulu region (Zambia). 1. Genera Micrasterias and Allorgeia (Desmidiales)’, *Plant Ecology and Evolution*, 147(3), pp. 392–404. Available at: https://doi.org/10.5091/plecevo.2014.985.

Commey, N.A., Magome, J. and Ishidaira, H. (2023) ‘Catchment-Scale Land Use and Land Cover Change Analysis’, pp. 1–21.

Congalton, R.G. (2001) ‘Accuracy assessment and validation of remotely sensed and other spatial information’, *International Journal of Wildland Fire*, 10(3–4), pp. 321–328. Available at: https://doi.org/10.1071/wf01031.

Convention on Wetlands (2022) *Briefing Note No. 13: Wetlands and agriculture: impacts of farming practices and pathways to sustainability*. Available at: http://www.fao.org/docrep/meeting/022/K8024E.pdf.

Corwin, D.L. and Yemoto, K. (2017) ‘Salinity: Electrical Conductivity and Total Dissolved Solids’, *Methods of Soil Analysis*, 2(1), p. 0. Available at: https://doi.org/10.2136/msa2015.0039.

Cunningham, P. and Delany, S.J. (2021) ‘K-Nearest Neighbour Classifiers-A Tutorial’, *ACM Computing Surveys*, 54(6). Available at: https://doi.org/10.1145/3459665.

Cutler, A., Cutler, D.R. and Stevens, J.R. (2012) ‘Ensemble Machine Learning’, *Ensemble*

Machine Learning [Preprint], (February 2014). Available at: https://doi.org/10.1007/978-1-4419-9326-7.

Dalianis, H. (2018) 'Evaluation Metrics and Evaluation', in *Clinical Text Mining*, pp. 45–53. Available at: https://doi.org/10.1007/978-3-319-78503-5_6.

Damm, C. (2022) 'Wetland Ecosystem Services', *Encyclopedia of Inland Waters*, (February), pp. 267–275. Available at: https://doi.org/10.1016/b978-0-12-819166-8.00154-7.

Davidson, N.C. (2014) 'How much wetland has the world lost? Long-term and recent trends in global wetland area', *Marine and Freshwater Research*, 65(10), pp. 934–941. Available at: https://doi.org/10.1071/MF14173.

Davidson, N.C. *et al.* (2019) 'A review of the adequacy of reporting to the Ramsar Convention on change in the ecological character of wetlands', *Marine and Freshwater Research*, 71(1), pp. 117–126. Available at: https://doi.org/10.1071/MF18328.

Delle Grazie, F.M. and Gill, L.W. (2022) 'Review of the Ecosystem Services of Temperate Wetlands and Their Valuation Tools', *Water (Switzerland)*, 14(9). Available at: https://doi.org/10.3390/w14091345.

Dey, J. and Vijay, R. (2021) 'A critical and intensive review on assessment of water quality parameters through geospatial techniques', *Environmental Science and Pollution Research*. Springer Science and Business Media Deutschland GmbH, pp. 41612–41626. Available at: https://doi.org/10.1007/s11356-021-14726-4.

Dey, N.N. *et al.* (2021) 'Geospatial modelling of changes in land use/land cover dynamics using Multi-layer perception Markov chain model in Rajshahi City, Bangladesh', *Environmental Challenges*, 4(March), p. 100148. Available at: https://doi.org/10.1016/j.envc.2021.100148.

Diggs, H.E. and Parker, J.M. (2009) 'Aquatic facilities', in *Planning and Designing Research Animal Facilities*. Elsevier Inc., pp. 323–331. Available at: https://doi.org/10.1016/B978-0-12-369517-8.00023-2.

Ding, X. *et al.* (2018) 'Electrical conductivity of nutrient solution influenced photosynthesis, quality, and antioxidant enzyme activity of pakchoi (Brassica campestris L. ssp. Chinensis) in a hydroponic system', *PLoS ONE*, 13(8). Available at: https://doi.org/10.1371/JOURNAL.PONE.0202090.

Dojlido, J., Raniszewski, J. and Woyciechowska, J. (1994) 'Water quality index - Application

for rivers in Vistula river basin in Poland', in *Water Science and Technology*. Kluwer Academic Publishers, pp. 57–64. Available at: https://doi.org/10.2166/wst.1994.0511.

Du, C.J. and Sun, D.W. (2008) 'Object Classification Methods', in *Computer Vision Technology for Food Quality Evaluation*. Academic Press, pp. 81–107. Available at: https://doi.org/10.1016/B978-012373642-0.50007-7.

Du, Y. *et al.* (2016) 'Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band', *Remote Sensing*, 8(4). Available at: https://doi.org/10.3390/rs8040354.

Dudley, N., Stolton, S. and Owusu, R.A. (2003) *Running pure : the importance of forest protected areas to drinking water : a research report for the World Bank/WWF Alliance for Forest Conservation and Sustainable Use*. Available at: https://documents1.worldbank.org/curated/en/414701468765561300/pdf/292830Running0pure.pdf (Accessed: 8 May 2024).

Dumitru, C. and Maria, V. (2013) 'Advantages and Disadvantages of Using Neural Networks for Predictions', in D.T. Epure and E.C. Pătariu (eds) *Ovidius University Annals, Series Economic Sciences* . Constanța: EBSCO Industries, Inc., pp. 444–449. Available at: http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203.

Ebrahimi-Khusfi, Z., Nafarzadegan, A.R. and Dargahian, F. (2021) 'Predicting the number of dusty days around the desert wetlands in southeastern Iran using feature selection and machine learning techniques', *Ecological Indicators*, 125(February), p. 107499. Available at: https://doi.org/10.1016/j.ecolind.2021.107499.

Ennaji, W. *et al.* (2018) 'Remote sensing approach to assess salt-affected soils in the north-east part of Tadla plain, Morocco', *Geology, Ecology, and Landscapes*, 2(1), pp. 22–28. Available at: https://doi.org/10.1080/24749508.2018.1438744.

Environmental Watch with Landsat satellites (2014). Available at: https://www.nasa.gov/content/goddard/environmental-watch-with-nasa-s-landsat-satellites/ (Accessed: 9 April 2022).

Etikan, I. (2016) 'Comparison of Convenience Sampling and Purposive Sampling', *American Journal of Theoretical and Applied Statistics*, 5(1), p. 1. Available at: https://doi.org/10.11648/j.ajtas.20160501.11.

Fanshawe, D. (1971) *The vegetation of Zambia,*. Lusaka: Printed by the Government Printer. Available at: https://www.worldcat.org/title/vegetation-of-zambia/oclc/1323510 (Accessed:

29 April 2022).

FAO (2017) ‘Water Pollution from Agriculture: A Global Review.’, *The Food and Agricultural Organization.*, pp. 1–35. Available at: https://www.fao.org/3/i7754e/i7754e.pdf (Accessed: 16 December 2023).

Farrier, D. and Tucker, L. (2000) ‘Wise use of wetlands under the Ramsar convention: A challenge for meaningful implementation of international law’, *Journal of Environmental Law*, 12(1), pp. 21–42. Available at: https://doi.org/10.1093/jel/12.1.21.

Fenta, A.A. *et al.* (2020) ‘Cropland expansion outweighs the monetary effect of declining natural vegetation on ecosystem services in sub-Saharan Africa’, *Ecosystem Services*, 45(April), p. 101154. Available at: https://doi.org/10.1016/j.ecoser.2020.101154.

Fernández-Delgado, M. *et al.* (2014) ‘Do we need hundreds of classifiers to solve real world classification problems?’, *Journal of Machine Learning Research*, 15(April 2020), pp. 3133–3181. Available at: https://doi.org/DOI:10.1117/1.JRS.11.015020.

Ferreira, A.J. and Figueiredo, M.A.T. (2012) ‘Boosting algorithms: A review of methods, theory, and applications’, *Ensemble Machine Learning: Methods and Applications*, pp. 35–85. Available at: https://doi.org/10.1007/9781441993267_2.

Foley, J.. *et al.* (2005) ‘Global Consequences of Land Use’, *Science*, 309, pp. 570–574. Available at: https://www.jstor.org/stable/3842335 (Accessed: 12 March 2022).

Foody, G. (2008) ‘Harshness in image classification accuracy assessment’, *International Journal of Remote Sensing*, 29(11), pp. 3137–3158. Available at: https://doi.org/10.1080/01431160701442120.

French, M.A. *et al.* (2021) ‘A planetary health model for reducing exposure to faecal contamination in urban informal settlements: Baseline findings from Makassar, Indonesia’, *Environment International*, 155, p. 106679. Available at: https://doi.org/10.1016/j.envint.2021.106679.

Gagné, S.A. and Fahrig, L. (2007) ‘Effect of landscape context on anuran communities in breeding ponds in the National Capital Region, Canada’, *Landscape Ecology*, 22(2), pp. 205–215. Available at: https://doi.org/10.1007/S10980-006-9012-3.

Galatowitsch, S.M. (2018) ‘Natural and Anthropogenic Drivers of Wetland Change’, in *The Wetland Book II: Distribution, Description, and Conservation*. Springer, Dordrecht, pp. 359–367. Available at: https://doi.org/10.1007/978-94-007-4001-3_217.

Gani, M.A. *et al.* (2023) ‘Assessing the impact of land use and land cover on river water quality using water quality index and remote sensing techniques’, *Environmental Monitoring and Assessment,* 195(4). Available at: https://doi.org/10.1007/s10661-023-10989-1.

Gardner, R.C. *et al.* (2015) ‘State of the World’s Wetlands and Their Services to People: A Compilation of Recent Analyses’, *SSRN Electronic Journal* [Preprint], (October 2017). Available at: https://doi.org/10.2139/ssrn.2589447.

Gardner, R.C., Okuno, E. and Pritchard, D. (2023) ‘Ramsar Convention governance and processes at the international level’, in *Ramsar Wetlands: Values, Assessment, Management*. Elsevier, pp. 37–67. Available at: https://doi.org/10.1016/B978-0-12-817803-4.00003-6.

Gerardo, R. and de Lima, I.P. (2022) ‘Sentinel-2 Satellite Imagery-Based Assessment of Soil Salinity in Irrigated Rice Fields in Portugal’, *Agriculture (Switzerland),* 12(9). Available at: https://doi.org/10.3390/agriculture12091490.

Gholami, R. and Fakhari, N. (2017) ‘Support Vector Machine: Principles, Parameters, and Applications’, in *Handbook of Neural Computation*. Academic Press, pp. 515–535. Available at: https://doi.org/10.1016/B978-0-12-811318-9.00027-2.

Gholizadeh, M.H., Melesse, A.M. and Reddi, L. (2016) ‘A comprehensive review on water quality parameters estimation using remote sensing techniques’, *Sensors (Switzerland).* Available at: https://doi.org/10.3390/s16081298.

Gislason, P.O., Benediktsson, J.A. and Sveinsson, J.R. (2006) ‘Random Forests for land cover classification’, *Pattern Recognition Letters*, 27(4), pp. 294–300. Available at: https://doi.org/10.1016/J.PATREC.2005.08.011.

Gómez, C., White, J.C. and Wulder, M.A. (2016) ‘Optical remotely sensed time series data for land cover classification: A review’, *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, pp. 55–72. Available at: https://doi.org/10.1016/J.ISPRSJPRS.2016.03.008.

Gorgoglione, A. *et al.* (2020) ‘Influence of land use/land cover on surface-water quality of Santa Lucia River, Uruguay’, *Sustainability (Switzerland),* 12(11). Available at: https://doi.org/10.3390/su12114692.

Grabowski, Z.J., Watson, E. and Chang, H. (2016) ‘Using spatially explicit indicators to investigate watershed characteristics and stream temperature relationships’, *Science of the Total Environment,* 551–552, pp. 376–386. Available at: https://doi.org/10.1016/j.scitotenv.2016.02.042.

Grimsdell, J. (1975) *Ecology of the black lechwe in the Bangweulu Basin of Zambia : Black Lechwe Research Project final report*. Lusaka Zambia: National Council for Scientific Research.

de Groot, D., Brander, L. and Max Finlayson, C. (2018) 'Wetland ecosystem services', *The Wetland Book: I: Structure and Function, Management, and Methods*, (Table 1), pp. 323–333. Available at: https://doi.org/10.1007/978-90-481-9659-3_66.

Guan, H. *et al.* (2013) 'Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests', *International Journal of Remote Sensing*, 34(14), pp. 5166–5186. Available at: https://doi.org/10.1080/01431161.2013.788261.

Gxokwe, S., Dube, T. and Mazvimavi, D. (2023) 'An assessment of long-term and large-scale wetlands change dynamics in the Limpopo transboundary river basin using cloud-based Earth observation data', *Wetlands Ecology and Management* [Preprint], (0123456789). Available at: https://doi.org/10.1007/s11273-023-09963-y.

Haji Gholizadeh, M., Melesse, A.M. and Reddi, L. (2016) 'Spaceborne and airborne sensors in water quality assessment', *International Journal of Remote Sensing*. Taylor & Francis, pp. 3143–3180. Available at: https://doi.org/10.1080/01431161.2016.1190477.

Hamdani, I. *et al.* (2018) 'Seasonal and diurnal evaporation from a deep hypersaline lake: The Dead Sea as a case study', *Journal of Hydrology*, 562(December 2017), pp. 155–167. Available at: https://doi.org/10.1016/j.jhydrol.2018.04.057.

Hamid, A., Bhat, S.U. and Jehangir, A. (2020) 'Local determinants influencing stream water quality', *Applied Water Science*, 10(1), pp. 1–16. Available at: https://doi.org/10.1007/s13201-019-1043-4.

Harrington, P. (2015) 'Machine Learning in Action', in *Efficient Learning Machines*, pp. 209–240. Available at: https://doi.org/10.1007/978-1-4302-5990-9_11.

Hayes, M.M., Miller, S.N. and Murphy, M.A. (2014) 'High-resolution landcover classification using random forest', *Remote Sensing Letters*, 5(2), pp. 112–121. Available at: https://doi.org/10.1080/2150704X.2014.882526.

Hemati, M. *et al.* (2021) 'A systematic review of landsat data for change detection applications: 50 years of monitoring the earth', *Remote Sensing*, 13(15). Available at: https://doi.org/10.3390/rs13152869.

Herbert, E.R. *et al.* (2015) 'A global perspective on wetland salinization: Ecological

consequences of a growing threat to freshwater wetlands', *Ecosphere*, 6(10). Available at: https://doi.org/10.1890/ES14-00534.1.

Hernández-Romero, G. *et al.* (2022) 'From Forest Dynamics to Wetland Siltation in Mountainous Landscapes: A RS-Based Framework for Enhancing Erosion Control', *Remote Sensing*, 14(8). Available at: https://doi.org/10.3390/rs14081864.

Hernandez, I. *et al.* (2020) 'Exploring sentinel-2 for land cover and crop mapping in portugal', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 43(B3), pp. 83–89. Available at: https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-83-2020.

Hestir, E.L. *et al.* (2015) 'Measuring freshwater aquatic ecosystems: The need for a hyperspectral global mapping satellite mission', *Remote Sensing of Environment*, 167, pp. 181–195. Available at: https://doi.org/10.1016/j.rse.2015.05.023.

Hintz, W.D. and Relyea, R.A. (2019) 'A review of the species, community, and ecosystem impacts of road salt salinisation in fresh waters', *Freshwater Biology*, 64(6), pp. 1081–1097. Available at: https://doi.org/10.1111/fwb.13286.

Holcomb, D.A. *et al.* (2020) 'Human fecal contamination of water, soil, and surfaces in households sharing poor-quality sanitation facilities in Maputo, Mozambique', *International Journal of Hygiene and Environmental Health*, 226(November 2019), p. 113496. Available at: https://doi.org/10.1016/j.ijheh.2020.113496.

Home | Ramsar Sites Information Service (no date). Available at: https://rsis.ramsar.org/?pagetab=1 (Accessed: 9 March 2022).

Horvat, M., Horvat, Z. and Pastor, K. (2021) 'Multivariate analysis of water quality parameters in Lake Palic, Serbia', *Environmental Monitoring and Assessment*, 193(7), pp. 1–18. Available at: https://doi.org/10.1007/s10661-021-09195-8.

Horwitz, P. (2022). Wetlands as social ecological systems, and relationality in the policy domain. Marine and Freshwater Research. https://doi.org/10.1071/MF22018

Hosseiny, B., Abdi, A.M. and Jamali, S. (2022) 'Urban land use and land cover classification with interpretable machine learning – A case study using Sentinel-2 and auxiliary data', *Remote Sensing Applications: Society and Environment*, 28(September), p. 100843. Available at: https://doi.org/10.1016/j.rsase.2022.100843.

Hu, Q. *et al.* (2019) 'Degradation of agricultural drainage water quantity and quality due to

farmland expansion and water-saving operations in arid basins', *Agricultural Water Management*, 213, pp. 185–192. Available at: https://doi.org/10.1016/j.agwat.2018.10.019.

Huang, G.Z. *et al.* (2020) 'Dilution and precipitation dominated regulation of stream water chemistry of a volcanic watershed', *Journal of Hydrology*, 583, p. 124564. Available at: https://doi.org/10.1016/j.jhydrol.2020.124564.

Hughes, R.H. and Hughes, J.S. (1992) 'A DIRECTORY OF AFRICAN WETLANDS'. Available at: https://doi.org/20.500.12592/25kcxb.

Hughes, S. (2019) 'Demystifying Theoretical and Conceptual Frameworks: A Guide for Students and Advisors of Educational Research', *Journal of Social Sciences*, 58(1–3), pp. 24–35. Available at: https://doi.org/10.31901/24566756.2019/58.1-3.2188.

Huntington, J.L. *et al.* (2017) 'Climate engine: Cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding', *Bulletin of the American Meteorological Society*, 98(11), pp. 2397–2409. Available at: https://doi.org/10.1175/BAMS-D-15-00324.1.

Hurley, T. and Mazumder, A. (2013) 'Spatial scale of land-use impacts on riverine drinking source water quality', *Water Resources Research*, 49(3), pp. 1591–1601. Available at: https://doi.org/10.1002/wrcr.20154.

Hyun, J.C. (2007) 'Effects of prevailing winds on turbidity of a shallow estuary', in *International Journal of Environmental Research and Public Health*. Int J Environ Res Public Health, pp. 185–192. Available at: https://doi.org/10.3390/ijerph2007040014.

Imandoust, S.B. and Bolandraftar, M. (2013) 'Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background', *Int. Journal of Engineering Research and Applications*, 3(5), pp. 605–610.

Islam, M.S., Azadi, M.A. and Nasiruddin, M. (2021) 'Water Quality of Boalia Khal Tributary of Halda River by Weighted Arithmetic Index Method', *American Journal of Water Resources*, 9(1), pp. 15–22. Available at: https://doi.org/10.12691/ajwr-9-1-3.

Jabareen, Y. (2009) 'Building a Conceptual Framework: Philosophy, Definitions, and Procedure', *International Journal of Qualitative Methods*, 8(4), pp. 49–62. Available at: https://doi.org/10.1177/160940690900800406.

Jeffery, R.C.. *et al.* (1986) *of Kafue Flats and Managing Basin Bangweulu*. Available at: chrome-

extension://efaidnbmnnnibpcajpcglclefindmkaj/https://portals.iucn.org/library/sites/library/files/documents/WTL-001-En.pdf.

Jordan, T.M. *et al.* (2023) 'Spatial structure of in situ reflectance in coastal and inland waters: implications for satellite validation', *Frontiers in Remote Sensing*, 4, p. 1249521. Available at: https://doi.org/10.3389/frsen.2023.1249521.

Jozkowski, A.C. (2017) 'Reason & Rigor: How Conceptual Frameworks Guide Research, 2nd Edition (2017)', *Occupational Therapy In Health Care*, 31(4), pp. 378–379. Available at: https://doi.org/10.1080/07380577.2017.1360538.

Kafy, A. Al *et al.* (2020) 'Modelling future land use land cover changes and their impacts on land surface temperatures in Rajshahi, Bangladesh', *Remote Sensing Applications: Society and Environment*, 18. Available at: https://doi.org/10.1016/j.rsase.2020.100314.

Kafy, A. Al *et al.* (2022) 'Predicting the impacts of land use/land cover changes on seasonal urban thermal characteristics using machine learning algorithms', *Building and Environment*, 217, p. 109066. Available at: https://doi.org/10.1016/j.buildenv.2022.109066.

Kafy, A. Al *et al.* (2023) 'Assessment and prediction of index based agricultural drought vulnerability using machine learning algorithms', *Science of the Total Environment*, 867. Available at: https://doi.org/10.1016/j.scitotenv.2023.161394.

Kafy, A.A. *et al.* (2021) 'Cellular Automata approach in dynamic modelling of land cover changes using RapidEye images in Dhaka, Bangladesh', *Environmental Challenges*, 4(April). Available at: https://doi.org/10.1016/j.envc.2021.100084.

Kalcheva, N., Todorova, M. and Marinova, G. (2020) 'NAIVE BAYES CLASSIFIER, DECISION TREE AND ADABOOST ENSEMBLE ALGORITHM – ADVANTAGES AND DISADVANTAGES', in *6th ERAZ Conference Proceedings (part of ERAZ conference collection)*. Association of Economists and Managers of the Balkans, Belgrade, Serbia, pp. 153–157. Available at: https://doi.org/10.31410/eraz.2020.153.

Kamble, V.H. and Dale, M.P. (2022) 'Machine learning approach for longitudinal face recognition of children', in *Machine Learning for Biometrics: Concepts, Algorithms and Applications*. Academic Press, pp. 1–27. Available at: https://doi.org/10.1016/B978-0-323-85209-8.00011-0.

Kamweneshe, B., Beilfuss, R. and Morrison, K. (2003) *Population and distribution of Wattled Cranes and other large waterbirds and large mammals on the Liuwa Plains National Park, Zambia*. Available at: https://doi.org/10.13140/RG.2.2.26628.32647.

Kang, H.Y., Rule, R.A. and Noble, P.A. (2012) 'Artificial Neural Network Modeling of Phytoplankton Blooms and its Application to Sampling Sites within the Same Estuary', in *Treatise on Estuarine and Coastal Science*. Elsevier Inc., pp. 161–172. Available at: https://doi.org/10.1016/B978-0-12-374711-2.00908-6.

Kaplan, G. and Avdan, U. (2017) 'Object-based water body extraction model using Sentinel-2 satellite imagery', *European Journal of Remote Sensing*, 50(1), pp. 137–143. Available at: https://doi.org/10.1080/22797254.2017.1297540.

Karakuş, C.B. (2020) 'Assessment of relationship between land use/cover and surface water quality trends within the riparian zone: A case study from sivas, turkey', *Desalination and Water Treatment*, 182, pp. 414–433. Available at: https://doi.org/10.5004/dwt.2020.25632.

Karmakar, S. and Mavukkandy, M.O. (2013) 'Lakes and reservoir : Pollution Encyclopedia of Environmental Management Lakes and Reservoirs : Pollution', *Encyclopedia of Environmental Management*, (May), pp. 1576–1587. Available at: https://doi.org/10.1081/E-EEM-120047215.

Khan, M., Ding, Q. and Perrizo, W. (2002) 'K-nearest neighbor classification on spatial data streams using P-trees', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2336, pp. 517–528. Available at: https://doi.org/10.1007/3-540-47887-6_51.

Khan, N.M. *et al.* (2005) 'Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators', in *Agricultural Water Management*. Elsevier, pp. 96–109. Available at: https://doi.org/10.1016/j.agwat.2004.09.038.

Khatri, N. and Tyagi, S. (2015) 'Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas', *Frontiers in Life Science*, 8(1), pp. 23–39. Available at: https://doi.org/10.1080/21553769.2014.933716.

King, R.S. *et al.* (2004) 'Watershed land use is strongly linked to PCBs in white perch in Chesapeake Bay subestuaries', *Environmental Science and Technology*, 38(24), pp. 6546–6552. Available at: https://doi.org/10.1021/es049059m.

King, R.S. *et al.* (2005) 'Spatial considerations for linking watershed land cover to ecological indicators in streams', *Ecological Applications*, 15(1), pp. 137–153. Available at: https://doi.org/10.1890/04-0481.

Kingsford, R.T. *et al.* (2021) 'Ramsar Wetlands of International Importance–Improving Conservation Outcomes', *Frontiers in Environmental Science*, 9. Available at:

https://doi.org/10.3389/fenvs.2021.643367.

Kivunja, C., & Kuyini, A. B. (2017). Understanding and Applying Research Paradigms in Educational Contexts. International Journal of Higher Education, 6(5), 26. https://doi.org/10.5430/ijhe.v6n5p26

Kolding, J. (2011) *A brief review of the Bangweulu fishery complex*. Available at: https://www.academia.edu/1155584/A_brief_review_of_the_Bangweulu_fishery_complex (Accessed: 10 June 2024).

Kulkarni, A.D. and Lowe, B. (2016) 'Random Forest Algorithm for Land Cover Classification', *International Journal on Recent and Innovation Trends in Computing and Communication*, 4(3), pp. 58–63. Available at: https://doi.org/http://hdl.handle.net/10950/341.

Kumar, A. and Kanaujia, A. (2018) 'Wetlands : Significance , Threats and their Conservation', *Envis center*, 7(March 2014), pp. 3&4. Available at: https://www.researchgate.net/publication/327816889.

Kumar, M. *et al.* (2024) 'In-situ optical water quality monitoring sensors—applications, challenges, and future opportunities', *Frontiers in Water*. Frontiers Media SA, p. 1380133. Available at: https://doi.org/10.3389/frwa.2024.1380133.

Kurtz, T.E. (2009) 'Regression and correlation.', in *Basic statistics.*, pp. 265–295. Available at: https://doi.org/10.1037/11782-009.

Laban, N. *et al.* (2019) 'Machine learning for enhancement land cover and crop types classification', *Studies in Computational Intelligence*, 801(April), pp. 71–87. Available at: https://doi.org/10.1007/978-3-030-02357-7_4.

Ladwig, R., Rock, L.A. and Dugan, H.A. (2023) 'Impact of salinization on lake stratification and spring mixing', *Limnology And Oceanography Letters*, 8(1), pp. 93–102. Available at: https://doi.org/10.1002/lol2.10215.

Lake Bangweulu Wetlands & Swamps (2021). Available at: https://www.zambiatourism.com/destinations/lakes/bangweulu/ (Accessed: 9 May 2022).

Landis, J.R. and Koch, G.G. (1977) 'The Measurement of Observer Agreement for Categorical Data', *Biometrics*, 33(1), p. 159. Available at: https://doi.org/10.2307/2529310.

Laonamsai, J. *et al.* (2023) 'Utilizing NDWI, MNDWI, SAVI, WRI, and AWEI for Estimating Erosion and Deposition in Ping River in Thailand', *Hydrology*, 10(3), pp. 1–25. Available at: https://doi.org/10.3390/hydrology10030070.

Lasanta, T. and Vicente-Serrano, S.M. (2012) ‘Complex land cover change processes in semiarid Mediterranean regions: An approach using Landsat images in northeast Spain’, *Remote Sensing of Environment*, 124, pp. 1–14. Available at: https://doi.org/10.1016/j.rse.2012.04.023.

Lehner, B. *et al.* (2021) ‘Identifying priority areas for surface water protection in data scarce regions: An integrated spatial analysis for Zambia’, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 31(8), pp. 1998–2016. Available at: https://doi.org/10.1002/aqc.3606.

Lei, C., Wagner, P.D. and Fohrer, N. (2022) ‘Influences of land use changes on the dynamics of water quantity and quality in the German lowland catchment of the Stör’, *Hydrology and Earth System Sciences*, 26(9), pp. 2561–2582. Available at: https://doi.org/10.5194/hess-26-2561-2022.

Leitão, P.J. *et al.* (2019) ‘A More Effective Ramsar Convention for the Conservation of Mediterranean Wetlands’, *Frontiers in Ecology and Evolution | www.frontiersin.org*, 7. Available at: https://doi.org/10.3389/fevo.2019.00021.

Leshem, S. and Trafford, V. (2007) ‘Overlooking the conceptual framework’, *Innovations in Education and Teaching International*. Routledge, pp. 93–105. Available at: https://doi.org/10.1080/14703290601081407.

Lestari, M.F. *et al.* (2023) ‘Analysis of mineral water quality based on SNI 3553:2015 and its consequences from legal perspectives’, in *IOP Conference Series: Earth and Environmental Science*. Available at: https://doi.org/10.1088/1755-1315/1190/1/012041.

Leyk, S. *et al.* (2018) *Assessing the accuracy of multi-temporal built-up land layers across rural-urban trajectories in the United States*, *Remote Sensing of Environment*. Available at: https://doi.org/10.1016/j.rse.2017.08.035.

Li, S. *et al.* (2008) ‘Water quality in relation to land use and land cover in the upper Han River Basin, China’, *Catena*, 75(2), pp. 216–222. Available at: https://doi.org/10.1016/j.catena.2008.06.005.

Li, S. *et al.* (2022) ‘Linking land use with riverine water quality: A multi-spatial scale analysis relating to various riparian strips’, *Frontiers in Environmental Science*, 10(November), pp. 1–15. Available at: https://doi.org/10.3389/fenvs.2022.1013318.

Li, Y.L. *et al.* (2012) ‘Relationship of land use/cover on water quality in the Liao River basin, China’, *Procedia Environmental Sciences*, 13, pp. 1484–1493. Available at:

https://doi.org/10.1016/j.proenv.2012.01.140.

Liberoff, A.L. *et al.* (2019) 'Assessing land use and land cover influence on surface water quality using a parametric weighted distance function', *Limnologica*, 74(October 2018), pp. 28–37. Available at: https://doi.org/10.1016/j.limno.2018.10.003.

Lihepanyama, D.L., Ndakidemi, P.A. and Treydte, A.C. (2022) 'Spatio–Temporal Water Quality Determines Algal Bloom Occurrence and Possibly Lesser Flamingo (Phoeniconaias minor) Presence in Momella Lakes, Tanzania', *Water (Switzerland)*, 14(21). Available at: https://doi.org/10.3390/w14213532.

Lintern, A. *et al.* (2018) 'Key factors influencing differences in stream water quality across space', *Wiley Interdisciplinary Reviews: Water*, 5(1), pp. 1–31. Available at: https://doi.org/10.1002/WAT2.1260.

Lizcano-Sandoval, L. *et al.* (2022) 'Seagrass distribution, areal cover, and changes (1990–2021) in coastal waters off West-Central Florida, USA', *Estuarine, Coastal and Shelf Science*, 279. Available at: https://doi.org/10.1016/j.ecss.2022.108134.

López-Calderón, J.M. and Riosmena-Rodríguez, R. (2016) 'Wetlands', in *Encyclopedia of Earth Sciences Series*. Springer, Dordrecht, pp. 738–741. Available at: https://doi.org/10.1007/978-94-017-8801-4_399.

Louppe, G. (2014) *Understanding Random Forests: From Theory to Practice*. Available at: http://arxiv.org/abs/1407.7502.

Loussaief, S. and Abdelkrim, A. (2017) 'Machine learning framework for image classification', in *2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, SETIT 2016*, pp. 58–61. Available at: https://doi.org/10.1109/SETIT.2016.7939841.

Lu, D. *et al.* (2004) 'Change detection techniques', *International Journal of Remote Sensing*, 25(12), pp. 2365–2401. Available at: https://doi.org/10.1080/0143116031000139863.

Luapula Provincial Administration » About Luapula (2022). Available at: https://www.lua.gov.zm/?page_id=4454 (Accessed: 1 April 2022).

Ma, L. *et al.* (2017) 'A review of supervised object-based land-cover image classification', *ISPRS Journal of Photogrammetry and Remote Sensing*. Elsevier, pp. 277–293. Available at: https://doi.org/10.1016/j.isprsjprs.2017.06.001.

Mahmoud, R. *et al.* (2023) 'Machine Learning-Based Land Use and Land Cover Mapping

Using Multi-Spectral Satellite Imagery: A Case Study in Egypt', *Sustainability (Switzerland)*, 15(12), pp. 1–21. Available at: https://doi.org/10.3390/su15129467.

Maitima, J.M. *et al.* (2010) 'The linkages between land use change, land degradation and biodiversity across East Africa', *African Journal of Environmental Science and Technology*, 3(10), pp. 310–325. Available at: https://doi.org/10.4314/ajest.v3i10.56259.

Marselina, M., Wibowo, F. and Mushfiroh, A. (2022) 'Water quality index assessment methods for surface water: A case study of the Citarum River in Indonesia', *Heliyon*, 8(7), p. e09848. Available at: https://doi.org/10.1016/j.heliyon.2022.e09848.

Maxwell, A.E., Warner, T.A. and Fang, F. (2018) 'Implementation of machine-learning classification in remote sensing: An applied review', *International Journal of Remote Sensing*, 39(9), pp. 2784–2817. Available at: https://doi.org/10.1080/01431161.2018.1433343.

Mayr, A. *et al.* (2014) 'The evolution of boosting algorithms: From machine learning to statistical modelling', *Methods of Information in Medicine*, 53(6), pp. 419–427. Available at: https://doi.org/10.3414/ME13-01-0122.

Mccann, K. (2017) 'Population and distribution of Wattled Cranes , Shoebills , and other large waterbirds in the Bangweulu Swamps , Zambia Population and distribution of Wattled Cranes , Shoebills , and other large waterbirds in the Bangweulu Swamps , Zambia', pp. 1–31. Available at: https://doi.org/10.13140/RG.2.2.26628.32647.

McDowell, R.W. (2021) 'Land use and water quality', *New Zealand Journal of Agricultural Research*, 64(3), pp. 269–270. Available at: https://doi.org/10.1080/00288233.2021.1933093.

McFeeters, S.K. (1996) 'The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features', *International Journal of Remote Sensing*, 17(7), pp. 1425–1432. Available at: https://doi.org/10.1080/01431169608948714.

McHugh, M.L. (2012) 'Interrater reliability: The kappa statistic', *Biochemia Medica*, 22(3), pp. 276–282. Available at: https://doi.org/10.11613/bm.2012.031.

Meena, S.R. *et al.* (2021) 'Chamoli disaster: pronounced changes in water quality and flood plains using Sentinel data', *Environmental Earth Sciences*, 80(17). Available at: https://doi.org/10.1007/s12665-021-09904-z.

Mello, K. de *et al.* (2020) 'Multiscale land use impacts on water quality: Assessment, planning, and future perspectives in Brazil', *Journal of Environmental Management*.

Academic Press. Available at: https://doi.org/10.1016/j.jenvman.2020.110879.

Meng, Z. *et al.* (2023) 'Post-2020 biodiversity framework challenged by cropland expansion in protected areas', *Nature Sustainability* [Preprint]. Available at: https://doi.org/10.1038/s41893-023-01093-w.

Miller, H.J. (2004) 'Tobler's first law and spatial analysis', *Annals of the Association of American Geographers*. Taylor & Francis Group, pp. 284–289. Available at: https://doi.org/10.1111/j.1467-8306.2004.09402005.x.

Ministry of Lands Natural Resources and Environmental Protection (2015) *Ministry of Lands Natural Resources and Environmental Protection: United Nations Convention on Biological Diversity: United Nations Convention on Biological Diversity Fifth National Report*. Available at: https://www.cbd.int/doc/world/zm/zm-nr-05-en.pdf (Accessed: 21 March 2022).

Ministry of Lands Natural Resources and Environmental and Protection (2015) 'Zambia's Second National Biodiversity Strategy and Action Plan (NBSAP-2)', 1(1), p. 80. Available at: https://www.cbd.int/doc/world/zm/zm-nbsap-v2-en.pdf.

Mohammadpour, P., Viegas, D.X. and Viegas, C. (2022) 'Vegetation Mapping with Random Forest Using Sentinel 2 and GLCM Texture Feature—A Case Study for Lousã Region, Portugal', *Remote Sensing*, 14(18). Available at: https://doi.org/10.3390/rs14184585.

Moraes, D. *et al.* (2021) 'INFLUENCE OF SAMPLE SIZE IN LAND COVER CLASSIFICATION ACCURACY USING RANDOM FOREST AND SENTINEL-2 DATA IN PORTUGAL', in *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4232–4235. Available at: https://doi.org/10.1109/IGARSS47720.2021.9553924.

Morales-Olmedo, M., Ortiz, M. and Sellés, G. (2015) 'Effects of transient soil waterlogging and its importance for rootstock selection', Chilean Journal of Agricultural Research. Instituto de Investigaciones Agropecuarias, INIA, pp. 45–56. Available at: https://doi.org/10.4067/S0718-58392015000300006.

Muche, M. *et al.* (2023) 'Land use and land cover changes and their impact on ecosystem service values in the north-eastern highlands of Ethiopia', *PLoS ONE*, 18(9 September). Available at: https://doi.org/10.1371/journal.pone.0289962.

Murray, C. *et al.* (2022) 'Water Quality Observations from Space: A Review of Critical Issues and Challenges', *Environments - MDPI*, 9(10). Available at: https://doi.org/10.3390/environments9100125.

Namugize, J.N., Jewitt, G. and Graham, M. (2018) 'Effects of land use and land cover changes on water quality in the uMngeni river catchment, South Africa', *Physics and*

Chemistry of the Earth, 105, pp. 247–264. Available at: https://doi.org/10.1016/j.pce.2018.03.013.

Nedd, R. *et al.* (2021) 'A Synthesis of Land Use/Land Cover Studies: Definitions, Classification Systems, Meta-Studies, Challenges and Knowledge Gaps on a Global Landscape', *Land*, 10(2020), pp. 1–30.

Ngoma, H. *et al.* (2017) *Irrigation Development for Climate Resilience in Zambia: The Known Knowns and Known Unknowns, Research gate*. Available at: https://www.canr.msu.edu/resources/irrigation-development-for-climate-resilience-in-zambia-the-known-knowns-and-known-unknowns (Accessed: 1 April 2022).

Nguvulu, A. *et al.* (2021) 'Surface Water Quality Response to Land Use Land Cover Change in an Urbanizing Catchment : A Case of Upper Chongwe River Catchment , Zambia', *Journal of Geographic Information System*, 13(5), pp. 578–602. Available at: https://doi.org/10.4236/JGIS.2021.135032.

Nguyen Van, H. *et al.* (2022) 'A comprehensive procedure to develop water quality index: A case study to the Huong river in Thua Thien Hue province, Central Vietnam', *PloS one*, 17(9), p. e0274673. Available at: https://doi.org/10.1371/journal.pone.0274673.

Nielsen, D.L. *et al.* (2003) 'Effects of increasing salinity on freshwater ecosystems in Australia', *Australian Journal of Botany*, 51(6), pp. 655–665. Available at: https://doi.org/10.1071/BT02115.

Niyoyitungiye, L., Giri, A. and Ndayisenga, M. (2020) 'Assessment of Coliforms Bacteria Contamination in Lake Tanganyika as Bioindicators of Recreational and Drinking Water Quality', *South Asian Journal of Research in Microbiology*, pp. 9–16. Available at: https://doi.org/10.9734/sajrm/2020/v6i330150.

Nkwanda, I.S. *et al.* (2021) 'Impact of land-use/land-cover dynamics on water quality in the Upper Lilongwe River basin, Malawi', *International Journal of Energy and Water Resources*, 5(2), pp. 193–204. Available at: https://doi.org/10.1007/s42108-021-00125-5.

Nogueira, G. *et al.* (2003) 'Microbiological quality of drinking water of urban and rural communities, Brazil', *Revista de Saude Publica*, 37(2), pp. 232–236. Available at: https://doi.org/10.1590/S0034-89102003000200011.

Ould Ahmed, B.A., Inoue, M. and Moritani, S. (2010) 'Effect of saline water irrigation and manure application on the available water content, soil salinity, and growth of wheat', *Agricultural Water Management*, 97(1), pp. 165–170. Available at:

https://doi.org/10.1016/j.agwat.2009.09.001.

Page, B.P., Olmanson, L.G. and Mishra, D.R. (2019) 'A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems', *Remote Sensing of Environment*, 231(June), p. 111284. Available at: https://doi.org/10.1016/j.rse.2019.111284.

Pal, M. *et al.* (2015) 'Electrical Conductivity of Lake Water as Environmental Monitoring – A Case Study of Rudrasagar Lake', *IOSR Journal of Environmental Science Ver. I*, 9(3), pp. 2319–2399. Available at: https://doi.org/10.9790/2402-09316671.

Pandey, S., Kumari, N. and Al Nawajish, S. (2023) 'Land Use Land Cover (LULC) and Surface Water Quality Assessment in and around Selected Dams of Jharkhand using Water Quality Index (WQI) and Geographic Information System (GIS)', *Journal of the Geological Society of India*, 99(2), pp. 205–218. Available at: https://doi.org/10.1007/s12594-023-2288-y.

Paulino, R.S. *et al.* (2022) 'Assessment of Adjacency Correction over Inland Waters Using Sentinel-2 MSI Images', *Remote Sensing*, 14(8). Available at: https://doi.org/10.3390/rs14081829.

Peng, K. *et al.* (2024) 'Detailed wetland-type classification using Landsat-8 time-series images: a pixel- and object-based algorithm with knowledge (POK)', *GIScience & Remote Sensing*, 61(1). Available at: https://doi.org/10.1080/15481603.2023.2293525.

Pereira, S. and Joshi, U. (2014) 'Implementation of Support Vector Machine Technique in Feedback Analysis System', *International Journal of Computer Applications*, 96(17), pp. 24–28. Available at: https://doi.org/10.5120/16887-6906.

Perennou, C. *et al.* (2020) 'Evolution of wetlands in Mediterranean region', *Water Resources in the Mediterranean Region*, pp. 297–320. Available at: https://doi.org/10.1016/B978-0-12-818086-0.00011-X.

Peterson, E.E. *et al.* (2011) 'A comparison of spatially explicit landscape representation methods and their relationship to stream condition', *Freshwater Biology*, 56(3), pp. 590–610. Available at: https://doi.org/10.1111/j.1365-2427.2010.02507.x.

Phethi, M.D. and Gumbo, J.R. (2019) 'Assessment of impact of land use change on the wetland in Makhitha village, Limpopo province, South Africa', *Jamba: Journal of Disaster Risk Studies*, 11(2). Available at: https://doi.org/10.4102/jamba.v11i2.693.

Pinto, N., Antunes, A.P. and Roca, J. (2021) ‘A cellular automata model for integrated simulation of land use and transport interactions’, *ISPRS International Journal of Geo-Information,* 10(3), pp. 1–16. Available at: https://doi.org/10.3390/ijgi10030149.

Pisner, D.A. and Schnyer, D.M. (2019) ‘Support vector machine’, in *Machine Learning: Methods and Applications to Brain Disorders*. Academic Press, pp. 101–121. Available at: https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

Piyathilake, I.D.U.H. *et al.* (2022) ‘Assessing groundwater quality using the Water Quality Index (WQI) and GIS in the Uva Province, Sri Lanka’, *Applied Water Science,* 12(4), pp. 1–19. Available at: https://doi.org/10.1007/s13201-022-01600-y.

Polikar, R. (2006) ‘Ensemble based systems in decision making’, *IEEE Circuits and Systems Magazine,* pp. 21–44. Available at: https://doi.org/10.1109/MCAS.2006.1688199.

Pooja, A. (2018) ‘Physical , Chemical and Biological Characteristics of Water (e Content Module) Paper Name : 5 Water Resources and Management Module : 26 Physical , Chemical and Biological Characteristics of Water Development Team’, *Water Resources and Management,* 1(5), pp. 2–17.

Powers, S.M. *et al.* (2023) ‘Spatially Intensive Patterns of Water Clarity in Reservoirs Determined Rapidly With Sensor-Equipped Boats and Satellites’, *Journal of Geophysical Research: Biogeosciences,* 128(10). Available at: https://doi.org/10.1029/2023JG007650.

Prajwala, T.. (2015) ‘A Comparative Study on Decision Tree and Random Forest Using R Tool’, *IJARCCE,* 4(1), pp. 196–199. Available at: https://doi.org/10.17148/ijarcce.2015.4142.

Qiu, C., Jiang, L. and Li, C. (2015) ‘Not always simple classification: Learning SuperParent for class probability estimation’, *Expert Systems with Applications,* 42(13), pp. 5433–5440. Available at: https://doi.org/10.1016/j.eswa.2015.02.049.

Quang, N.H. *et al.* (2023) ‘Calibration of Sentinel-2 Surface Reflectance for Water Quality Modelling in Binh Dinh’s Coastal Zone of Vietnam’, *Sustainability (Switzerland)* , 15(2). Available at: https://doi.org/10.3390/su15021410.

Rahaman, Z.A. *et al.* (2022) ‘Assessing the impacts of vegetation cover loss on surface temperature, urban heat island and carbon emission in Penang city, Malaysia’, *Building and Environment,* 222, p. 109335. Available at: https://doi.org/10.1016/j.buildenv.2022.109335.

Ramadas, M. and Samantaray, A.K. (2018) ‘Applications of Remote Sensing and GIS in Water Quality Monitoring and Remediation: A State-of-the-Art Review’, *Energy,*

Environment, and Sustainability, pp. 225–246. Available at: https://doi.org/10.1007/978-981-10-7551-3_13/COVER.

Ramezan, C.A. *et al.* (2021) 'Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data', *Remote Sensing*, 13(3), pp. 1–27. Available at: https://doi.org/10.3390/rs13030368.

Ramsar Convention on Wetlands (2018) *Global Wetland Outlook: State of the World's Wetlands and their Services to People*. Gland, Switzerland. Available at: https://www.researchgate.net/publication/328093181_Global_Wetland_Outlook_State_of_the_World's_Wetlands_and_Their_Services_to_People.

Rather, I.A. and Dar, A.Q. (2020) 'Assessing the impact of land use and land cover dynamics on water quality of Dal Lake, NW Himalaya, India', *Applied Water Science*, 10(10), pp. 1–18. Available at: https://doi.org/10.1007/s13201-020-01300-5.

Ren, J.-P. *et al.* (2021) 'Gold enrichment characteristics and exploration prospects in Zambia: Based on 1:1000000 geochemical mapping'. Available at: https://doi.org/10.1016/S2096-5192(22)00086-6.

Rey-Romero, D.C., Domínguez, I. and Oviedo-Ocaña, E.R. (2022) 'Effect of agricultural activities on surface water quality from páramo ecosystems', *Environmental Science and Pollution Research*, 29(55), pp. 83169–83190. Available at: https://doi.org/10.1007/s11356-022-21709-6.

Rodriguez-Galiano, V.F. *et al.* (2012) 'An assessment of the effectiveness of a random forest classifier for land-cover classification', *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), pp. 93–104. Available at: https://doi.org/10.1016/j.isprsjprs.2011.11.002.

Rodríguez-López, L. *et al.* (2021) 'Retrieving water turbidity in araucanian lakes (South-central chile) based on multispectral landsat imagery', *Remote Sensing*, 13(16), pp. 1–16. Available at: https://doi.org/10.3390/rs13163133.

Rokach, L. (2010) 'Ensemble-based classifiers', *Artificial Intelligence Review*, 33(1–2), pp. 1–39. Available at: https://doi.org/10.1007/s10462-009-9124-7.

Rokach, L. and Maimon, O. (2006) 'Decision Trees', in *Data Mining and Knowledge Discovery Handbook*. New York: Springer-Verlag, pp. 165–192. Available at: https://doi.org/10.1007/0-387-25465-x_9.

Romero, L.S., Marcello, J. and Vilaplana, V. (2020) 'Super-resolution of Sentinel-2 imagery

using generative adversarial networks', *Remote Sensing*, 12(15). Available at: https://doi.org/10.3390/RS12152424.

Roy, K., Kar, S. and Das, R.N. (2015) 'Selected Statistical Methods in QSAR', in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Academic Press, pp. 191–229. Available at: https://doi.org/10.1016/b978-0-12-801505-6.00006-5.

Russ, J. *et al.* (2022) 'The impact of water quality of GDP growth: Evidence from around the world', *Water Security*. Elsevier, p. 100130. Available at: https://doi.org/10.1016/j.wasec.2022.100130.

Rusydi, A.F. (2018) 'Correlation between conductivity and total dissolved solid in various type of water: A review', *IOP Conference Series: Earth and Environmental Science*, 118(1). Available at: https://doi.org/10.1088/1755-1315/118/1/012019.

Rwanga, S.S. and Ndambuki, J.M. (2017) 'Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS', *International Journal of Geosciences*, 08(04), pp. 611–622. Available at: https://doi.org/10.4236/ijg.2017.84033.

Sadiq, R., Rodriguez, M.J. and Mian, H.R. (2019) 'Empirical models to predict disinfection by-products (DBPs) in drinking water: An updated review', in *Encyclopedia of Environmental Health*. Elsevier, pp. 324–338. Available at: https://doi.org/10.1016/B978-0-12-409548-9.11193-5.

Sallam, G.A.H. and Elsayed, E.A. (2018) 'Estimating relations between temperature, relative humidity as independed variables and selected water quality parameters in Lake Manzala, Egypt', *Ain Shams Engineering Journal*, 9(1), pp. 1–14. Available at: https://doi.org/10.1016/j.asej.2015.10.002.

Sarp, G. and Ozcelik, M. (2017) 'Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey', *Journal of Taibah University for Science*, 11(3), pp. 381–391. Available at: https://doi.org/10.1016/j.jtusci.2016.04.005.

Schapire, R.E. *et al.* (1999) 'Improved Boosting Algorithms Using Confidence-rated Predictions', 37, pp. 297–336.

Semeniuk, C.A. and Semeniuk, V. (1995) 'A geomorphic approach to global classification for inland wetlands', *Vegetatio*, 118(1–2), pp. 103–124. Available at: https://doi.org/10.1007/BF00045193.

Sen, S. and Janssen, J. (1989) *REPORT PREPARED FOR THE AQUACULTURE FOR LOCAL COMMUNITY DEVELOPMENT PROGRAMME*. Available at: https://www.fao.org/3/ac987e/AC987E09.htm (Accessed: 1 April 2022).

Senthilkumar, M. (2010) 'Use of artificial neural networks (ANNs) in colour measurement', in *Colour Measurement: Principles, Advances and Industrial Applications*. Elsevier Ltd, pp. 125–146. Available at: https://doi.org/10.1533/9780857090195.1.125.

Shadrin, N., Stetsiuk, A. and Anufriieva, E. (2022) 'Differences in Mercury Concentrations in Water and Hydrobionts of the Crimean Saline Lakes: Does Only Salinity Matter?', *Water (Switzerland)*, 14(17). Available at: https://doi.org/10.3390/w14172613.

Shafizadeh-Moghadam, H. (2019) 'Improving spatial accuracy of urban growth simulation models using ensemble forecasting approaches', *Computers, Environment and Urban Systems*, 76, pp. 91–100. Available at: https://doi.org/10.1016/j.compenvurbsys.2019.04.005.

Sharma, S. and Bhattacharya, A. (2017) 'Drinking water contamination and treatment techniques', *Applied Water Science*, 7(3), pp. 1043–1067. Available at: https://doi.org/10.1007/s13201-016-0455-7.

Van Sickle, J. and Burch Johnson, C. (2008) 'Parametric distance weighting of landscape influence on streams', *Landscape Ecology*, 23(4), pp. 427–438. Available at: https://doi.org/10.1007/s10980-008-9200-4.

Simaika, J.P., Chakona, A. and van Dam, A.A. (2021) 'Editorial: Towards the Sustainable Use of African Wetlands', *Frontiers in Environmental Science*, 9(March), pp. 1–4. Available at: https://doi.org/10.3389/fenvs.2021.658871.

Singh, N., Kaur, M. and Katnoria, J.K. (2017) 'Analysis on bioaccumulation of metals in aquatic environment of Beas River Basin: A case study from Kanjli wetland', *GeoHealth*, 1(3), pp. 93–105. Available at: https://doi.org/10.1002/2017GH000062.

Smith, L.C. *et al.* (2007) 'Rising minimum daily flows in northern Eurasian rivers: A growing influence of groundwater in the high-latitude hydrologic cycle', *Journal of Geophysical Research: Biogeosciences*, 112(4). Available at: https://doi.org/10.1029/2006JG000327.

Soeprobowati, T.R. *et al.* (2021) 'Physico-chemical and biological water quality of Warna and Pengilon Lakes, Dieng, Central Java', *Journal of Water and Land Development*, 51, pp. 38–49. Available at: https://doi.org/10.24425/jwld.2021.139013.

Song, Y. *et al.* (2020) 'Effects of land use on stream water quality in the rapidly urbanized

areas: A multiscale analysis', *Water (Switzerland)*, 12(4). Available at: https://doi.org/10.3390/W12041123.

Sputh, B.H.C. *et al.* (2008) 'Integration of in-situ and remote sensing data for water risk management', in *Proc. iEMSs 4th Biennial Meeting - Int. Congress on Environmental Modelling and Software: Integrating Sciences and Information Technology for Environmental Assessment and Decision Making, iEMSs 2008*, pp. 486–493. Available at: https://www.academia.edu/51335230/Integration_of_In_situ_and_remote_sensing_Data_for_Water_Risk_Management (Accessed: 26 February 2024).

Stephenson, P.J., Ntiamoa-Baidu, Y. and Simaika, J.P. (2020) 'The Use of Traditional and Modern Tools for Monitoring Wetlands Biodiversity in Africa: Challenges and Opportunities', *Frontiers in Environmental Science*, 8, p. 61. Available at: https://doi.org/10.3389/FENVS.2020.00061/BIBTEX.

Storrs, J. (1995) *Know your trees: Some of the common trees found in Zambia. Regional Soil Conservation Unit, (RSCU).* Nairobi: Regional Soil Conservation Unit. Available at: https://www.worldcat.org/title/know-your-trees-some-of-the-common-trees-found-in-zambia/oclc/7979762/editions?referer=di&editionsView=true (Accessed: 29 April 2022).

Strauss, M. (2017) 'Flotilla of tiny satellites will photograph the entire Earth every day', *Science* [Preprint]. Available at: https://doi.org/10.1126/SCIENCE.AAL0811.

Stuip, M., Baker, C. and Oosterberg, W. (2002) 'The socio-economics of wetlands', *Wetlands International and RIZA*, p. 40. Available at: http://elibrary.cenn.org/Wetlands/The Socio-Economics of Wetlands.pdf.

Suchan, J. and Azam, S. (2021) 'Effect of salinity on evaporation from water surface in bench-scale testing', *Water (Switzerland)*, 13(15). Available at: https://doi.org/10.3390/w13152067.

Sun, Y., Wong, A.K.C. and Kamel, M.S. (2009) 'Classification of imbalanced data: A review', *International Journal of Pattern Recognition and Artificial Intelligence*, 23(4), pp. 687–719. Available at: https://doi.org/10.1142/S0218001409007326.

Swamee, P.K. and Tyagi, A. (2007) 'Improved Method for Aggregation of Water Quality Subindices', *Journal of Environmental Engineering*, 133(2), pp. 220–225. Available at: https://doi.org/10.1061/(asce)0733-9372(2007)133:2(220).

Syeed, M.M.M. *et al.* (2023) 'Surface water quality profiling using the water quality index, pollution index and statistical methods: A critical review', *Environmental and Sustainability*

Indicators, 18(March), p. 100247. Available at: https://doi.org/10.1016/j.indic.2023.100247.

Szabó, S., Gácsi, Z. and Balázs, B. (2016) 'Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories', *Landscape & Environment*, 10(3–4), pp. 194–202. Available at: https://doi.org/10.21120/le/10/3-4/13.

Szczerbicki, E. (2001) 'Management of Complexity and Information Flow', in *Agile Manufacturing: The 21st Century Competitive Strategy*. Elsevier Science Ltd, pp. 247–263. Available at: https://doi.org/10.1016/b978-008043567-1/50013-9.

Tahiru, A.A., Doke, D.A. and Baatuuwie, B.N. (2020a) 'Effect of land use and land cover changes on water quality in the Nawuni Catchment of the White Volta Basin, Northern Region, Ghana', *Applied Water Science*, 10(8), pp. 1–14. Available at: https://doi.org/10.1007/s13201-020-01272-6.

Tahiru, A.A., Doke, D.A. and Baatuuwie, B.N. (2020b) 'Effect of land use and land cover changes on water quality in the Nawuni Catchment of the White Volta Basin, Northern Region, Ghana', *Applied Water Science*, 10(8), pp. 1–14. Available at: https://doi.org/10.1007/s13201-020-01272-6.

Taiwo, B.E. *et al.* (2023) 'Monitoring and predicting the influences of land use/land cover change on cropland characteristics and drought severity using remote sensing techniques', *Environmental and Sustainability Indicators*, 18(March), p. 100248. Available at: https://doi.org/10.1016/j.indic.2023.100248.

Talukdar, S. *et al.* (2020) 'Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations — A Review'. Available at: https://doi.org/https://doi.org/10.3390/rs12071135.

Taylor, M., Elliott, H.A. and Navitsky, L.O. (2018) 'Relationship between total dissolved solids and electrical conductivity in Marcellus hydraulic fracturing fluids', *Water Science and Technology*, 77(8), pp. 1998–2004. Available at: https://doi.org/10.2166/wst.2018.092.

Tehrany, M.S., Pradhan, B. and Jebuv, M.N. (2014) 'A comparative assessment between object and pixel-based classification approaches for land use/land cover mapping using SPOT 5 imagery', *Geocarto International*, 29(4), pp. 351–369. Available at: https://doi.org/10.1080/10106049.2013.768300.

Thakur, R. and Panse, P. (2022) 'Classification Performance of Land Use from Multispectral Remote Sensing Images using Decision Tree, K-Nearest Neighbor, Random Forest and Support Vector Machine Using EuroSAT Data', *International Journal of Intelligent Systems*

and Applications in Engineering, 10(1s), pp. 67–77. Available at: https://ijisae.org/index.php/IJISAE/article/view/2238 (Accessed: 7 July 2023).

Thamaga, K.H., Dube, T. and Shoko, C. (2022) 'Evaluating the impact of land use and land cover change on unprotected wetland ecosystems in the arid-tropical areas of South Africa using the Landsat dataset and support vector machine', *Geocarto International*, 37(25), pp. 10344–10365. Available at: https://doi.org/10.1080/10106049.2022.2034986.

Thorp, J.H. and Covich, A.P. (2015) 'Overview of Inland Water Habitats', in *Thorp and Covich's Freshwater Invertebrates: Ecology and General Biology: Fourth Edition*. Academic Press, pp. 23–56. Available at: https://doi.org/10.1016/B978-0-12-385026-3.00002-4.

Tu, J. V. (1996) 'Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes', *Journal of Clinical Epidemiology*, 49(11), pp. 1225–1231. Available at: https://doi.org/10.1016/S0895-4356(96)00002-9.

Turunen, K. *et al.* (2020) 'Analysing Contaminant Mixing and Dilution in River Waters Influenced by Mine Water Discharges', *Water, Air, and Soil Pollution*, 231(6). Available at: https://doi.org/10.1007/s11270-020-04683-y.

Ullman, J.L. (2013) 'Agricultural systems: The basics', *Agricultural & Biological Engineering*, pp. 1–36.

United Nations Environment Programme (2008) 'Water Quality for Ecosystem and Human Health', *Taiwan Review*. Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.cbd.int/doc/health/health-waterquality-en.pdf.

United States Geological Survey (2013) 'Landsat 8 band designations'. Available at: https://www.usgs.gov/media/images/landsat-8-band-designations (Accessed: 1 May 2022).

US EPA (2001) 'Types of Wetands', *Encyclopedia of Ecology*, pp. 1690–1697.

USGS (2022) *What is the Landsat satellite program and why is it important? | U.S. Geological Survey, https://www.usgs.gov/faqs/what-landsat-satellite-program-and-why-it-important#:~:text=Landsat%20satellites%20have%20the%20optimal%20ground%20resolution%20and,a%20host%20of%20other%20natural%20and%20human-caused%20changes.* Available at: https://www.usgs.gov/faqs/what-landsat-satellite-program-and-why-it-important (Accessed: 9 April 2022).

Vanhellemont, Q. (2019) 'Adaptation of the dark spectrum fitting atmospheric correction for

aquatic applications of the Landsat and Sentinel-2 archives', *Remote Sensing of Environment*, 225(February), pp. 175–192. Available at: https://doi.org/10.1016/j.rse.2019.03.010.

Vogt, W. and Johnson, R. (2015) *Correlation and Regression Analysis, Correlation and Regression Analysis*. Available at: https://doi.org/10.4135/9781446286104.

Walton, N.R.G. (1989) 'Electrical Conductivity and Total Dissolved Solids—What is Their Precise Relationship?', *Desalination*, 72(3), pp. 275–292. Available at: https://doi.org/10.1016/0011-9164(89)80012-8.

Wambugu, N. *et al.* (2021) 'A hybrid deep convolutional neural network for accurate land cover classification', *International Journal of Applied Earth Observation and Geoinformation*, 103. Available at: https://doi.org/10.1016/j.jag.2021.102515.

Wang, L.M. *et al.* (2006) 'Combining decision tree and Naive Bayes for classification', *Knowledge-Based Systems*, 19(7), pp. 511–515. Available at: https://doi.org/10.1016/j.knosys.2005.10.013.

Wang, W. *et al.* (2023) 'Long-Term Changes in Water Body Area Dynamic and Driving Factors in the Middle-Lower Yangtze Plain Based on Multi-Source Remote Sensing Data', *Remote Sensing*, 15(7), p. 1816. Available at: https://doi.org/10.3390/rs15071816.

Wang, X., Zhang, F. and Ding, J. (2017) 'Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed, China', *Scientific Reports*, 7(1), pp. 1–18. Available at: https://doi.org/10.1038/s41598-017-12853-y.

Wang, Y. *et al.* (2021) 'Relating land-use/land-cover patterns to water quality in watersheds based on the structural equation modeling', *Catena*, 206, p. 105566. Available at: https://doi.org/10.1016/j.catena.2021.105566.

Wang, Z. *et al.* (2020) 'Water level decline in a reservoir: Implications for water quality variation and pollution source identification', *International Journal of Environmental Research and Public Health*, 17(7). Available at: https://doi.org/10.3390/ijerph17072400.

Warren, E.L. *et al.* (2019) 'Heavy Metals, Iron, and Arsenic in Water and Sediment from a Cold Spring in Southwest Ohio', *Environmental Engineering Science*, 36(10), pp. 1296–1306. Available at: https://doi.org/10.1089/ees.2019.0177.

Wetland degradation, a big threat to water management in Zambia — Abertine Rift Conservation Society (2017). Available at: http://www.arcosnetwork.org/en/article/stakeholders-meet-for-days-in-kigali-for-scenario-

guided-review-and-regional-policy-harmonization-workshop-26 (Accessed: 26 March 2022).

Wetlands for Sustainable Cities | WWF Zambia (2018). Available at: https://www.wwfzm.panda.org/?27825/Wetlands-for-Sustainable-Cities (Accessed: 12 April 2022).

Winton, R.S. *et al.* (2021) 'Anthropogenic influences on Zambian water quality: Hydropower and land-use change', *Environmental Science: Processes and Impacts*, 23(7), pp. 981–994. Available at: https://doi.org/10.1039/d1em00006c.

Wolpert, D.H. and Macready, W.G. (1997) 'No free lunch theorems for optimization', *IEEE Transactions on Evolutionary Computation*, 1(1), pp. 67–82. Available at: https://doi.org/10.1109/4235.585893.

Woodcock, C.E. *et al.* (2008) 'Free access to landsat imagery', *Science*. American Association for the Advancement of Science, p. 1011. Available at: https://doi.org/10.1126/science.320.5879.1011a.

Wu, X. *et al.* (2022) 'Application of Phosphate Materials as Constructed Wetland Fillers for Efficient Removal of Heavy Metals from Wastewater', *International Journal of Environmental Research and Public Health*, 19(9), p. 5344. Available at: https://doi.org/10.3390/ijerph19095344.

Wulder, M.A. *et al.* (2018) 'Land cover 2.0', *International Journal of Remote Sensing*, 39(12), pp. 4254–4284. Available at: https://doi.org/10.1080/01431161.2018.1452075.

Wulder, M.A. *et al.* (2019) 'Current status of Landsat program, science, and applications', *Remote Sensing of Environment*, 225, pp. 127–147. Available at: https://doi.org/10.1016/j.rse.2019.02.015.

WWF (2020) *World's wetlands disappearing three times faster than forests | WWF*. Available at: https://wwf.panda.org/wwf_news/?335575/Worlds-wetlands-disappearing-three-times-faster-than-forests (Accessed: 26 March 2022).

Xu, H. (2006a) 'Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery', *International Journal of Remote Sensing*, 27(14), pp. 3025–3033. Available at: https://doi.org/10.1080/01431160600589179.

Xu, H. (2006b) 'Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery', *International Journal of Remote Sensing*, 27(14), pp. 3025–3033. Available at: https://doi.org/10.1080/01431160600589179.

Xu, T. *et al.* (2019) 'Wetlands of international importance: Status, threats, and future protection', *International Journal of Environmental Research and Public Health*, 16(10). Available at: https://doi.org/10.3390/ijerph16101818.

Yang, H. *et al.* (2022) 'A Review of Remote Sensing for Water Quality Retrieval : Progress and Challenges'.

Yang, J. *et al.* (2021) 'Understanding land surface temperature impact factors based on local climate zones', *Sustainable Cities and Society*, 69(March), p. 102818. Available at: https://doi.org/10.1016/j.scs.2021.102818.

Yuan, T. *et al.* (2019) 'Urbanization impacts the physicochemical characteristics and abundance of fecal markers and bacterial pathogens in surface water', *International Journal of Environmental Research and Public Health*, 16(10). Available at: https://doi.org/10.3390/ijerph16101739.

Yuh, Y.G. *et al.* (2023) 'Application of machine learning approaches for land cover monitoring in northern Cameroon', *Ecological Informatics*, 74(December 2022), p. 101955. Available at: https://doi.org/10.1016/j.ecoinf.2022.101955.

Zainurin, S.N. *et al.* (2022) 'Advancements in Monitoring Water Quality Based on Various Sensing Methods: A Systematic Review', *International Journal of Environmental Research and Public Health*, 19(21). Available at: https://doi.org/10.3390/ijerph192114080.

Zambia Environment Outlook Report 4 (2017). Available at: www.zema.org.zm.

Zambia Homepage — Lake Tanganyika (2011). Available at: http://lta.iwlearn.org/Countries/z (Accessed: 31 March 2022).

Zambia Wildlife Authority (2006) *Information Sheet on Ramsar Wetlands (RIS)– 2006-2008 version*. Available at: https://rsis.ramsar.org/RISapp/files/RISrep/ZM531RIS.pdf.

Zedler, J.B. and Kercher, S. (2005) 'Wetland resources: Status, trends, ecosystem services, and restorability', *Annual Review of Environment and Resources*, pp. 39–74. Available at: https://doi.org/10.1146/annurev.energy.30.050504.144248.

Zhang, F.F. *et al.* (2011) 'Comparative analysis of automatic water identification method based on multispectral remote sensing', in *Procedia Environmental Sciences*. Elsevier, pp. 1482–1487. Available at: https://doi.org/10.1016/j.proenv.2011.12.223.

Zhang, M. *et al.* (2021a) 'Simulating the relationship between land use/cover change and urban thermal environment using machine learning algorithms in Wuhan city, China', *Land*,

11(1). Available at: https://doi.org/10.3390/land11010014.

Zhang, M. *et al.* (2021b) 'Simulating the Relationship between Land Use/Cover Change and Urban Thermal Environment Using Machine Learning Algorithms in Wuhan City, China', *Land 2022, Vol. 11, Page 14*, 11(1), p. 14. Available at: https://doi.org/10.3390/LAND11010014.

Zhang, X. *et al.* (2020) 'Spatiotemporal variability and key influencing factors of river fecal coliform within a typical complex watershed', *Water Research*, 178, p. 115835. Available at: https://doi.org/10.1016/j.watres.2020.115835.

Zhang, Y. *et al.* (2020) 'Impacts of land use changes on wetland ecosystem services in the tumen river basin', *Sustainability (Switzerland)*, 12(23), pp. 1–15. Available at: https://doi.org/10.3390/su12239821.

Zhang, Y. and Thorburn, P.J. (2022) 'Handling missing data in near real-time environmental monitoring: A system and a review of selected methods', *Future Generation Computer Systems*, 128, pp. 63–72. Available at: https://doi.org/10.1016/j.future.2021.09.033.

Zhou, W. *et al.* (2006) 'Mapping the concentrations of total suspended matter in Lake Taihu, China, using Landsat-5 TM data', *International Journal of Remote Sensing*, 27(6), pp. 1177–1191. Available at: https://doi.org/10.1080/01431160500353825.

Zhou, X. *et al.* (2023) 'An Overview of Coastline Extraction from Remote Sensing Data', *Remote Sensing*. Multidisciplinary Digital Publishing Institute, p. 4865. Available at: https://doi.org/10.3390/rs15194865.

Zhu, Z. *et al.* (2019) 'Benefits of the free and open Landsat data policy', *Remote Sensing of Environment*, 224, pp. 382–385. Available at: https://doi.org/10.1016/J.RSE.2019.02.016.

Zorrilla-Miras, P. *et al.* (2014) 'Effects of land-use change on wetland ecosystem services: A case study in the Doñana marshes (SW Spain)', *Landscape and Urban Planning*, 122(February), pp. 160–174. Available at: https://doi.org/10.1016/j.landurbplan.2013.09.013.

Zou, J., Han, Y. and So, S.S. (2008) 'Overview of artificial neural networks', *Methods in Molecular Biology*, pp. 15–23. Available at: https://doi.org/10.1007/978-1-60327-101-1_2.

APPENDICES

Appendix I: The Ethical Clearance Letter

THE UNIVERSITY OF ZAMBIA
DIRECTORATE OF RESEARCH AND GRADUATE STUDIES

Great East Road Campus | P.O. Box 32379 | Lusaka10101 | Tel: +260-211-290 258/291 777 **Fax**: (+260)-211-290 258/253 952 | E-**mail**: director.drgs@unza.zm | Website: www.unza.zm

APPROVAL OF STUDY

IORG No. 0005376
NASREC IRB No. 00006465

15th September, 2022

REF NO. NASREC-2022-JUL-026

Misheck Lesa Chundu
The University of Zambia
School of Mines
P.O. Box 32379
LUSAKA

Dear Mr. Chundu,

RE: "THE EFFECTS OF LAND-USED CHANGE ON ECOSYSTEM SERVICES AND THE FUTURE SCENARIO OF THE BANGWEULU WETLANDS, ZAMBIA: A REMOTE SENSING APPROACH"

Reference is made to your submission of the protocol captioned above. The NASREC resolved to approve this study and your participation as Principal Investigator for a period of one year.

REVIEW TYPE	**ORDINARY REVIEW**	**APPROVAL NO. NASREC-2022-JUL-026**
Approval and Expiry Date	Approval Date: 15th September, 2022	Expiry Date: 14th September, 2023
Protocol Version and Date	Version - Nil.	14th September, 2023
Information Sheet, Consent Forms and Dates	□ English.	To be provided
Consent form ID and Date	Version - Nil	To be provided
Recruitment Materials	Nil	Nil
Other Study Documents	Questionnaire.	
Number of Participants Approved for Study		

Specific conditions will apply to this approval. As Principal Investigator it is your responsibility to ensure that the contents of this letter are adhered to. If these are not adhered to, the approval may be suspended. Should the study be suspended, study sponsors and other regulatory authorities will be informed.

CONDITIONS OF APPROVAL

- No participant may be involved in any study procedure prior to the study approval or after the expiration date.
- All unanticipated or Serious Adverse Events (SAEs) must be reported to NASREC within 5 days.
- All protocol modifications must be approved by NASREC prior to implementation unless they are intended to reduce risk (but must still be reported for approval). Modifications will include any change of investigator/s or site address.
- All protocol deviations must be reported to NASREC within 5 working days.
- All recruitment materials must be approved by NASREC prior to being used.
- Principal investigators are responsible for initiating Continuing Review proceedings. HSSREC will only approve a study for a period of 12 months.
- It is the responsibility of the PI to renew his/her ethics approval through a renewal application to NASREC.
- Where the PI desires to extend the study after expiry of the study period, documents for study extension must be received by NASREC at least 30 days before the expiry date. This is for the purpose of facilitating the review process. Documents received within 30 days after expiry will be labelled "late submissions" and will incur a penalty fee of K500.00. No study shall be renewed whose documents are submitted for renewal 30 days after expiry of the certificate.
- Every 6 (six) months a progress report form supplied by The University of Zambia Natural and Applied Sciences Research Ethics Committee as an IRB must be filled in and submitted to us. There is a penalty of K500.00 for failure to submit the report.
- When closing a project, the PI is responsible for notifying, in writing or using the Research Ethics and Management Online (REMO), both NASREC and the National Health Research Authority (NHRA) when ethics certification is no longer required for a project.
- In order to close an approved study, a Closing Report must be submitted in writing or through the REMO system. A Closing Report should be filed when data collection has ended and the study team will no longer be using human participants or animals or secondary data or have any direct or indirect contact with the research participants or animals for the study.
- Filing a closing report (rather than just letting your approval lapse) is important as it assists NASREC in efficiently tracking and reporting on projects. Note that some funding agencies and sponsors require a notice of closure from the IRB which had approved the study and can only be generated after the Closing Report has been filed.

- A reprint of this letter shall be done at a fee.

- All protocol modifications must be approved by NASREC by way of an application for an amendment prior to implementation unless they are intended to reduce risk (but must still be reported for approval). Modifications will include any change of investigator/s or site address or methodology and methods. Many modifications entail minimal risk adjustments to a protocol and/or consent form and can be made on an Expedited basis (via the IRB Chair). Some examples are: format changes, correcting spelling errors, adding key personnel, minor changes to questionnaires, recruiting and changes, and so forth. Other, more substantive changes, especially those that may alter the risk-benefit ratio, may require Full Board review. In all cases, except where noted above regarding subject safety, any changes to any protocol document or procedure must first be approved by NASREC before they can be implemented.

Should you have any questions regarding anything indicated in this letter, please do not hesitate to get in touch with us at the above indicated address.

On behalf of NASREC, we would like to wish you all the success as you carry out your study. Yours

faithfully,

Dr.E. M. Mwanaumo

DR. E. M. MWANAUMO

CHAIRPERSON
THE UNIVERSITY OF ZAMBIA NATURAL AND APPLIED SCIENCES RESEARCH ETHICS COMMITTEE - IRB

CC: Director, Directorate of Research and Graduate Studies
Assistant Director (Research), Directorate of Research and Graduate Studies Assistant
Registrar (Research), Directorate of Research and Graduate Studies

Appendix II: Field Geometry points of LULC classes used in Image Classification

S/N	Class	X	Y	S/N	Class	X	Y
1	Built_up	29.643786	-11.1132	43	Built_up	29.550799	-11.34024
2	Built_up	29.624714	-11.1432	44	Built_up	29.557647	-11.342733
3	Built_up	29.602481	-11.1408	45	Built_up	29.567707	-11.321322
4	Built_up	29.555734	-11.1662	46	Built_up	29.524247	-11.269342
5	Built_up	29.623876	-11.1332	47	Built_up	29.545653	-11.366393
6	Built_up	29.657139	-11.0815	48	Built_up	29.561355	-11.38547
7	Built_up	28.928106	-11.1571	49	Built_up	29.566435	-11.393225
8	Built_up	30.624457	-12.916	50	Built_up	30.225191	-13.225556
9	Built_up	30.612307	-12.952	51	Built_up	30.218325	-13.230392
10	Built_up	30.60501	-12.9466	52	Built_up	30.241451	-13.230207
11	Built_up	30.233452	-13.2298	53	Built_up	29.94721	-11.114362
12	Built_up	30.410463	-11.9966	54	Built_up	29.986306	-11.104928
13	Built_up	30.384923	-11.9834	55	Built_up	30.116689	-11.208741
14	Built_up	29.574662	-12.1691	56	Built_up	29.814821	-11.210248
15	Built_up	29.557735	-12.1631	57	Built_up	28.872097	-11.188409
16	Built_up	28.892144	-11.1653	58	Built_up	28.870124	-11.207908
17	Built_up	28.885342	-11.126	59	Built_up	28.556758	-11.136596
18	Built_up	28.889833	-11.1134	60	Built_up	28.737062	-11.925104
19	Built_up	28.512658	-11.2333	61	Cropland	29.051996	-11.869066
20	Built_up	28.516583	-11.2226	62	Cropland	30.245478	-12.413608
21	Built_up	28.897384	-11.0798	63	Cropland	29.43295	-12.332702
22	Built_up	29.928972	-10.2625	64	Cropland	30.090828	-11.212417
23	Built_up	29.914376	-10.248	65	Cropland	30.017044	-11.146697
24	Built_up	29.994678	-12.2959	66	Cropland	30.150467	-13.226875
25	Built_up	30.025537	-12.3001	67	Cropland	30.162789	-13.207518
26	Built_up	29.269469	-11.3928	68	Cropland	30.129611	-13.214517
27	Built_up	29.277538	-11.4049	69	Cropland	30.060499	-13.177254
28	Built_up	29.409576	-11.3059	70	Cropland	30.053487	-13.175566
29	Built_up	30.691131	-12.8862	71	Cropland	30.068105	-13.160045
30	Built_up	30.637041	-12.9107	72	Cropland	30.061488	-13.185018
31	Built_up	30.617614	-12.9249	73	Cropland	30.061976	-13.163244
32	Built_up	28.903303	-11.2051	74	Cropland	30.05526	-13.162189
33	Built_up	28.875752	-11.2285	75	Cropland	30.081323	-13.168239
34	Built_up	28.880073	-11.2338	76	Cropland	30.074717	-13.171719
35	Built_up	28.904909	-11.2167	77	Cropland	30.067928	-13.170956
36	Built_up	28.892434	-11.163	78	Cropland	30.049541	-13.177212
37	Built_up	28.868798	-11.1753	79	Cropland	29.888631	-13.092575
38	Built_up	28.929232	-11.2074	80	Cropland	29.88318	-13.083886
39	Built_up	29.061403	-11.2631	81	Cropland	29.928223	-13.072376
40	Built_up	29.110699	-11.2724	82	Cropland	29.947623	-13.096289
41	Built_up	29.109094	-11.2064	83	Cropland	29.91347	-13.109719
42	Built_up	29.110032	-11.2121	84	Cropland	29.902978	-13.116618

S/N	Class	X	Y		S/N	Class	X	Y
85	Forest	30.125264	-9.988838		127	Water	29.7779	-12.149731
86	Forest	30.12816	-10.322038		128	Water	29.8327	-12.135407
87	Forest	30.106194	-10.337515		129	Water	29.8449	-12.111435
88	Forest	30.064857	-10.320565		130	Water	29.7918	-12.046697
89	Forest	30.061957	-10.308764		131	Water	29.8055	-11.982389
90	Forest	29.875562	-10.168844		132	Water	29.7718	-11.967429
91	Forest	29.872588	-10.17303		133	Water	29.7988	-11.937741
92	Forest	29.767012	-10.369655		134	Water	29.8971	-11.795526
93	Forest	29.781776	-10.405335		135	Water	29.7288	-10.741274
94	Forest	29.636795	-10.41764		136	Water	29.6446	-10.718805
95	Forest	29.623803	-10.446221		137	Water	29.6387	-10.738125
96	Forest	29.424815	-10.62915		138	Water	30.5187	-11.410646
97	Forest	29.020589	-12.201114		139	Water	29.6881	-11.140093
98	Forest	29.236066	-12.285763		140	Water	29.8645	-11.098323
99	Forest	29.301741	-12.258884		141	Water	29.8445	-11.104089
100	Forest	29.318121	-12.260029		142	Water	29.8295	-11.120379
101	Forest	30.840655	-12.57246		143	Water	29.7872	-11.011301
102	Forest	30.084659	-10.461207		144	Water	29.8363	-10.965737
103	Forest	30.078543	-10.451592		145	Water	29.6408	-11.269499
104	Forest	30.064802	-10.451596		146	Water	29.6994	-11.277363
105	Forest	30.099784	-10.376403		147	Water	29.6943	-11.326151
106	Forest	30.106614	-10.388977		148	Water	29.6749	-11.392934
107	Forest	30.095071	-10.321497		149	Water	29.7638	-11.342184
108	Forest	30.271303	-10.39505		150	Water	29.8715	-11.198621
109	Forest	30.32801	-10.367592		151	Water	29.7393	-10.947716
110	Forest	28.842392	-11.82061		152	Water	29.7109	-10.932747
111	Forest	30.071299	-12.891581		153	Water	29.6844	-11.514346
112	Forest	30.129983	-12.981664		154	Water	29.5549	-11.451983
113	Forest	29.837025	-13.412961		155	Water	29.7009	-11.737817
114	Forest	29.821263	-12.962222		156	Water	29.7173	-11.743531
115	Forest	29.850218	-12.508645		157	Water	29.8536	-11.683179
116	Forest	29.243119	-10.927331		158	Water	29.8664	-11.716378
117	Forest	29.159765	-10.978305		159	Water	29.9079	-11.696874
118	Forest	28.612474	-11.22663		160	Water	29.9151	-11.699008
119	Forest	28.576263	-11.307453		161	Water	29.5211	-11.213777
120	Forest	28.458143	-11.441064		162	Water	29.5174	-11.189203
121	Forest	28.446646	-11.587695		163	Water	29.6044	-11.200071
122	Forest	28.416322	-11.585519		164	Water	29.565	-11.239324
123	Forest	28.496552	-11.640792		165	Grass	29.8064	-11.749191
124	Forest	28.47437	-11.723241		166	Grass	29.8212	-11.757399
125	Forest	28.455516	-11.719134		167	Grass	29.8603	-11.754978
126	Forest	28.473644	-11.743008		168	Grass	29.8809	-11.734291

S/N	Class	X	Y	S/N	Class	X	Y
169	Cropland	30.306605	-12.7422	210	Cropland	29.79735	-10.655971
170	Cropland	30.294673	-12.7432	211	Cropland	29.728964	-10.657207
171	Cropland	30.322052	-12.7737	212	Cropland	29.63625	-10.85434
172	Cropland	30.304547	-12.7762	213	Cropland	28.570956	-11.065814
173	Cropland	30.310452	-12.7814	214	Forest	28.62103	-11.717569
174	Cropland	30.300834	-12.7737	215	Forest	29.049043	-12.013964
175	Cropland	30.381472	-12.4954	216	Forest	30.182588	-12.490366
176	Cropland	30.369792	-12.4529	217	Forest	30.19434	-12.496579
177	Cropland	30.377819	-12.4506	218	Forest	30.196582	-12.439104
178	Cropland	30.366188	-12.4683	219	Forest	30.282614	-12.477596
179	Cropland	30.361424	-12.4687	220	Forest	30.058339	-12.642931
180	Cropland	30.382257	-12.4677	221	Forest	30.054655	-12.636811
181	Cropland	30.390707	-12.4699	222	Forest	30.053131	-12.646603
182	Cropland	30.405119	-12.4744	223	Forest	29.88755	-13.169085
183	Cropland	30.374374	-12.496	224	Forest	29.893755	-13.165128
184	Cropland	28.980649	-11.1422	225	Forest	29.898577	-13.164376
185	Cropland	28.994502	-11.1292	226	Forest	29.886196	-13.175165
186	Cropland	29.002515	-11.1314	227	Forest	29.889447	-13.166453
187	Cropland	28.879074	-10.9654	228	Forest	29.898388	-13.154845
188	Cropland	28.871744	-10.9639	229	Forest	29.854948	-13.165305
189	Cropland	28.87446	-10.9724	230	Forest	29.867664	-13.161486
190	Cropland	28.779039	-11.0049	231	Forest	29.370935	-12.296772
191	Cropland	28.652327	-11.8637	232	Forest	29.426019	-12.334268
192	Cropland	28.656846	-11.8717	233	Forest	29.45219	-12.412178
193	Cropland	28.662841	-11.8642	234	Forest	29.457498	-12.406069
194	Cropland	28.652608	-11.8803	235	Forest	29.428036	-12.285649
195	Cropland	30.043992	-10.3579	236	Forest	29.439607	-12.25986
196	Cropland	30.196332	-13.2148	237	Forest	29.590153	-12.12515
197	Cropland	29.547743	-11.2997	238	Forest	29.596982	-12.124025
198	Cropland	29.111251	-11.7252	239	Forest	29.594688	-12.127775
199	Cropland	30.135994	-13.0774	240	Forest	29.545329	-12.00973
200	Cropland	30.193826	-13.1466	241	Forest	29.565237	-11.791969
201	Cropland	30.151341	-13.1917	242	Forest	29.45005	-11.774018
202	Cropland	30.089717	-13.2316	243	Forest	29.325432	-11.815003
203	Cropland	29.911624	-13.2621	244	Forest	29.302643	-11.399283
204	Cropland	29.904815	-13.2507	245	Forest	29.306681	-11.384096
205	Cropland	29.929917	-13.2253	246	Forest	29.303252	-11.376946
206	Cropland	29.918085	-13.3152	247	Forest	29.298544	-11.371541
207	Forest	29.200051	-11.3415	248	Forest	29.573207	-10.984565
208	Forest	29.231022	-11.2446	249	Forest	29.518966	-10.968989
209	Forest	29.581345	-10.9983	250	Forest	29.516426	-10.963399

S/N	Class	X	Y
251	Cropland	29.89449	-13.30981
252	Cropland	29.894316	-13.326495
253	Cropland	29.915661	-13.296061
254	Cropland	29.915869	-13.281529
255	Cropland	29.916172	-13.165595
256	Cropland	29.911975	-13.140205
257	Cropland	29.816411	-10.642111
258	Water	29.645219	-11.13452
259	Water	29.725933	-11.023426
260	Water	29.693249	-11.053589
261	Water	29.653656	-11.028827
262	Water	29.478785	-11.20329
263	Forest	29.472517	-10.976498
264	Forest	29.546127	-11.007344
265	Built_up	29.129263	-11.209332
266	Built_up	29.489181	-11.352183
267	Built_up	29.491179	-11.359546
268	Cropland	29.898833	-13.083315
269	Cropland	29.890473	-12.923011
270	Cropland	30.309271	-12.766861
271	Forest	28.515846	-11.692362
272	Forest	28.584038	-11.722014
273	Forest	28.618214	-11.698434
274	Grass	29.908708	-11.727212
275	Grass	29.93023	-11.710834
276	Grass	29.987814	-11.704798
277	Forest	29.231685	-11.311856
278	Grass	29.815114	-11.656837
279	Grass	30.583919	-10.876009
280	Grass	30.148923	-12.18641
281	Grass	30.215641	-12.128276
282	Grass	29.815043	-11.572695
283	Grass	29.860255	-11.569814
284	Grass	29.933035	-11.563551
285	Grass	30.067603	-12.242732
286	Grass	29.832734	-12.255128
287	Grass	29.887094	-12.2607
288	Grass	29.895323	-11.72312
289	Grass	30.036056	-12.080791
290	Grass	29.976197	-12.008868
291	Grass	29.897842	-11.992303
292	Grass	29.8135	-12.024864
293	Grass	29.886	-11.868383
294	Grass	29.8921	-11.880841
295	Grass	29.924	-11.847417
296	Grass	29.9369	-11.86075
297	Grass	29.8692	-11.951332
298	Grass	29.8751	-11.926163
299	Grass	29.9628	-11.904095
300	Grass	29.6951	-12.045265
301	Grass	29.6479	-12.07528
302	Grass	29.7102	-12.104413
303	Grass	29.7989	-12.115559
304	Grass	29.4972	-11.222578
305	Grass	29.4925	-11.240607
306	Grass	29.5581	-11.269567
307	Grass	29.8861	-12.110545
308	Grass	29.9771	-11.36352
309	Grass	30.1561	-12.100364
310	Grass	30.0328	-12.011652
311	Grass	30.0007	-12.007868
312	Grass	29.9741	-12.051009
313	Grass	29.9543	-12.09127
314	Grass	29.9903	-12.10812
315	Grass	30.0466	-12.072378

Appendix III: Geometry points for water quality sampling points used for water quality monitoring of open water bodies.

Sample ID	X	Y	EC	TDS	Turbidity	TH	Ca-H	Feacal	Temp	pH	Cl2+	Na-	Fe2+
DMC 34	29.72419	-11.062117	26.6	17	0.36	12	6	14	NIL	NIL	10	2.2	<0.002
DMC 59	29.736511	-11.429075	23	15	0.46	12	9	22	NIL	NIL	NIL	NIL	<0.002
DMC 33	29.681049	-11.085831	24.2	15	0.47	12	5	21	NIL	NIL	9	2.0	<0.002
DMC 24	29.695511	-11.075739	24.7	16	0.54	12	8	0	NIL	NIL	8	1.7	<0.002
DMC 22	29.72348	-11.051793	26.3	17	0.57	12	4	22	NIL	NIL	12	2.6	<0.002
DMC 64	29.621945	-11.399482	27.1	17	0.58	16	8	20	NIL	NIL	9	2.0	<0.002
DMC 56	29.801033	-11.386112	24	15	0.59	12	7	0	NIL	NIL	8	1.7	<0.002
DMC 21	29.721206	-11.057127	25.9	17	0.59	14	7	40	NIL	NIL	11	2.4	<0.002
DMC 40	29.627053	-11.437551	26.3	17	0.6	14	6	0	NIL	NIL	7	0.0	<0.002
DMC 44	29.728704	-11.440199	26.5	17	0.61	14	8	0	NIL	NIL	NIL	NIL	<0.002
DMC 32	29.651148	-11.116869	24.2	15	0.64	10	9	0	NIL	NIL	NIL	NIL	<0.002
DMC 25	29.67476	-11.102385	23.5	15	0.64	12	9	0	NIL	NIL	NIL	NIL	<0.002
DMC 63	29.64131	-11.410156	26.8	17	0.65	15	8	21	NIL	NIL	NIL	NIL	<0.002
DMC 30	29.613296	-11.168424	22.6	14	0.68	14	8	0	NIL	NIL	8	1.7	<0.002
DMC 50	29.8814	-11.365105	27.1	17	0.73	15	10	0	NIL	NIL	8	1.7	<0.002
DMC 9	29.708214	-11.091533	26	16	0.73	12	6	0	NIL	NIL	13	2.8	<0.002
DMC 58	29.756072	-11.409745	24.7	15	0.76	12	8	20	NIL	NIL	8	1.7	<0.002
DMC 28	29.62471	-11.165338	22.6	14	0.77	16	10	0	NIL	NIL	11	2.4	<0.002
DMC 12	29.771678	-11.0048	26.9	17	0.77	14	8	10	NIL	NIL	8	1.7	<0.002
DMC 60	29.705448	-11.446904	25.1	16	0.78	12	7	16	NIL	NIL	9	2.0	<0.002
DMC 38	29.579486	-11.408238	25.8	17	0.8	14	4	0	NIL	NIL	8	1.7	<0.002
DMC 36	29.568065	-11.366702	26.3	17	0.83	16	4	0	NIL	NIL	8	1.7	<0.002
DMC 27	29.64047	-11.148053	22.7	14	0.84	12	9	0	NIL	NIL	8	1.7	<0.002
DMC 15	29.80189	-10.948387	27.2	17	0.84	12	9	22	NIL	NIL	10	2.2	<0.002
DMC 51	29.893057	-11.344114	27.7	18	0.86	15	11	0	NIL	NIL	10	2.2	<0.002
DMC 62	29.66048	-11.424393	27.2	17	0.87	15	9	0	NIL	NIL	8	1.7	<0.002
DMC 48	29.827742	-11.407789	24.5	16	0.89	16	9	71	NIL	NIL	7	1.5	<0.002
DMC 41	29.654851	-11.443641	26.4	17	0.92	15	7	0	NIL	NIL	NIL	NIL	<0.002
DMC 55	29.822303	-11.375328	25.9	17	0.93	14	8	0	NIL	NIL	10	2.2	<0.002
DMC 31	29.627932	-11.152848	22.7	14	0.93	10	8	0	NIL	NIL	11	2.4	<0.002

Sample ID	X	Y	EC	TDS	Turbidity	TH	Ca-H	Feacal	Temp	pH	Cl2+	Na-	Fe2+
DMC 47	29.800745	-11.409802	25.1	16	0.95	12	11	26	NIL	NIL	11	2.4	<0.002
DMC 54	29.852962	-11.355684	27.3	18	0.96	14	9	0	NIL	NIL	7	1.5	<0.002
DMC 43	29.709268	-11.451119	26.3	17	0.97	17	8	22	NIL	NIL	8	1.7	<0.002
DMC 26	29.656403	-11.128306	23.3	15	0.97	12	9	0	NIL	NIL	11	2.4	<0.002
DMC 17	29.823527	-10.92309	27.3	17	0.98	12	8	21	NIL	NIL	NIL	NIL	<0.002
DMC 3	29.584303	-11.338946	27	17	1	14	8	0	NIL	NIL	10	2.2	<0.002
DMC 45	29.754866	-11.428527	24.8	16	1.01	12	9	0	NIL	NIL	11	2.4	<0.002
DMC 20	29.780938	-10.997988	27	17	1.01	14	7	16	NIL	NIL	7	1.5	<0.002
DMC 61	29.676449	-11.432518	27.5	17	1.04	18	9	14	NIL	NIL	7	1.5	<0.002
DMC 46	29.779035	-11.411643	24.5	16	1.04	12	10	14	NIL	NIL	8	1.7	<0.002
DMC 2	29.570557	-11.338889	27	17	1.04	10	5	0	NIL	NIL	8	1.7	<0.002
DMC 23	29.710146	-11.058755	25	16	1.05	18	4	16	NIL	NIL	7	1.5	<0.002
DMC 7	29.633386	-11.202608	27	17	1.08	12	9	24	NIL	NIL	10	2.2	<0.002
DMC 5	29.6052	-11.269622	26	17	1.12	14	8	36	NIL	NIL	8	1.7	<0.002
DMC 13	29.780244	-10.987452	27.1	17	1.12	10	9	12	NIL	NIL	9	2.0	<0.002
DMC 39	29.597868	-11.42258	25.9	17	1.13	14	4	0	NIL	NIL	10	2.2	<0.002
DMC 37	29.575684	-11.384999	26.5	18	1.17	14	5	0	NIL	NIL	8	1.7	<0.002
DMC 14	29.79209	-10.96691	27.1	17	1.17	14	8	50	NIL	NIL	9	2.0	<0.002
DMC 18	29.802727	-10.952206	27.9	18	1.21	14	9	24	NIL	NIL	NIL	NIL	<0.002
DMC 4	29.592294	-11.303567	27	17	1.23	16	8	41	NIL	NIL	NIL	NIL	<0.002
DMC 29	29.601476	-11.182116	22	14	1.23	12	9	0	NIL	NIL	11	2.4	<0.002
DMC 52	29.89833	-11.321882	26	17	1.31	14	10	23	NIL	NIL	7	1.5	<0.002
DMC 57	29.777488	-11.399812	24.8	16	1.32	12	7	49	NIL	NIL	9	2.0	<0.002
DMC 10	29.735255	-11.058767	27	17	1.32	12	6	9	NIL	NIL	10	2.2	<0.002
DMC 42	29.68707	-11.355635	26.5	17	1.39	15	9	18	NIL	NIL	7	1.5	<0.002
MC 19HM	29.581376	-11.175625	26	16	1.39	NIL	NIL	0	25	7.19	15	3.3	0.031
R2 5N	29.797233	-11.20653	26	17	1.4	NIL	NIL	95	25.1	7.31	NIL	NIL	<0.002
DMC 19	29.79488	-10.973546	27.8	18	1.41	8	3	20	NIL	NIL	NIL	NIL	<0.002
DMC 8	29.658186	-11.171544	27	17	1.52	14	7	0	NIL	NIL	12	2.6	<0.002

Sample ID	X	Y	EC	TDS	Turbidity	TH	Ca-H	Feacal	Temp	pH	Cl2+	Na-	Fe2+
DMC 16	29.815608	-10.925541	27.4	18	1.53	14	8	18	NIL	NIL	11	2.4	<0.002
DMC 49	29.849579	-11.3839	26.9	17	1.65	15	9	0	NIL	NIL	NIL	NIL	<0.002
DMC 11	29.758769	-11.032295	27	17	1.65	12	7	10	NIL	NIL	9	2.0	<0.002
R2 18 N	29.92355	-11.094194	26	17	1.74	NIL	NIL	0	24.5	7.73	NIL	NIL	<0.002
R2 11HM	29.945861	-11.22121	24	16	1.8	NIL	NIL	>100	24.4	6.92	NIL	NIL	0.226
MC 16 HM	29.554034	-11.332372	18	12	1.83	NIL	NIL	0	24.5	6.82	NIL	NIL	0.436
R2 1 HM	29.719498	-11.353044	28	18	1.9	NIL	NIL	0	25	6.71	NIL	NIL	<0.002
R2 4 N	29.812306	-11.197201	26	17	2.1	NIL	NIL	0	25.1	7.17	NIL	NIL	<0.002
R2 3N	29.800003	-11.220907	27	17	2.4	NIL	NIL	>100	24.9	7.14	NIL	NIL	<0.002
R2 21N	29.916852	-11.106046	27	17	2.41	NIL	NIL	20	26.5	7.63	NIL	NIL	<0.002
R2 6N	29.836231	-11.182676	27	17	2.6	NIL	NIL	85	26.6	7.36	NIL	NIL	<0.002
R2 7N	29.871275	-11.166722	27	17	2.67	NIL	NIL	>100	26.6	7.36	NIL	NIL	<0.002
DMC 1	29.56261	-11.355846	25	17	2.78	14	6	0	NIL	NIL	NIL	NIL	<0.002
MC 20 HM	29.619828	-11.158091	26	17	2.84	NIL	NIL	0	26.9	7.26	10	2.2	0.076
R2 11N	29.945861	-11.22121	24	16	2.9	NIL	NIL	>100	24.4	6.92	NIL	NIL	0.226
R2 17 N	29.940713	-11.109404	25	16	3	NIL	NIL	85	24.9	7.43	NIL	NIL	<0.002
R2 27N	29.625499	-11.301558	27	17	3.11	NIL	NIL	0	26	7.8	NIL	NIL	<0.002
R226 N	29.713177	-11.23442	26	17	3.2	NIL	NIL	>100	26.1	7.3	NIL	NIL	<0.002
R2 9N	29.890276	-11.190404	27.3	18	3.2	NIL	NIL	70	25.3	7.71	NIL	NIL	<0.002
R2 8N	29.874805	-11.174547	27.4	18	3.2	NIL	NIL	>100	25.4	7.62	NIL	NIL	<0.002
DMC 6	29.605383	-11.250623	26	17	3.24	14	7	38	NIL	NIL	10	2.2	<0.002
R2 25N	29.781044	-11.164617	26	17	3.4	NIL	NIL	0	25.8	7.67	NIL	NIL	<0.002
R2 22N	29.916852	-11.106046	26	17	3.49	NIL	NIL	0	24.7	7.76	NIL	NIL	<0.002
R2 10 N	29.946123	-11.232489	28	18	3.5	NIL	NIL	50	25.1	7.05	NIL	NIL	<0.002
R2 9HM	29.890276	-11.190404	27.5	18	3.7	NIL	NIL	70	25.3	7.71	NIL	NIL	<0.002
R2 16 HM	29.9028	-11.13499	27	17	4.62	NIL	NIL	>100	26.8	8.08	NIL	NIL	<0.002
R2 12HM	29.947686	-11.220309	24	16	4.7	NIL	NIL	>100	23.3	6.98	NIL	NIL	0.053
R223 N	29.876993	-11.097428	27	17	4.7	NIL	NIL	62	26.7	7.81	NIL	NIL	<0.002
R220 HM	29.845727	-11.071024	27	17	5.1	NIL	NIL	58	26.6	8.03	NIL	NIL	<0.002
MC 1 HM	29.751152	-11.46645	NIL	NIL	6.4	NIL	NIL	0	NIL	NIL	NIL	NIL	0.582

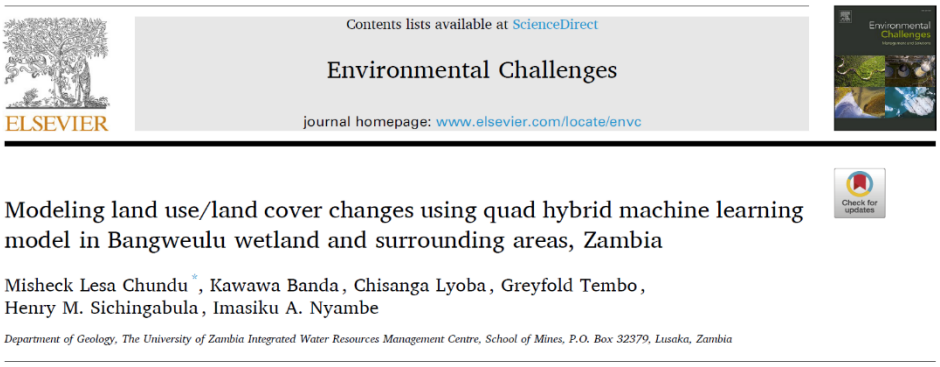
Sample ID	X	Y	EC	TDS	Turbidity	TH	Ca-H	Feacal	Temp	pH	Cl2+	Na-	Fe2+
MC 12HM	29.570389	-11.325786	NIL	NIL	6.52	NIL	NIL	0	NIL	NIL	NIL	NIL	0.085
DMC 53	29.881648	-11.334656	25.8	17	6.61	14	9	0	NIL	NIL	NIL	NIL	<0.002
R2 19N	29.888735	-11.067144	27	17	6.8	NIL	NIL	78	25	7.65	NIL	NIL	0.034
R2 24 N	29.813305	-11.128368	27	17	6.9	NIL	NIL	75	27	7.86	NIL	NIL	<0.002
DMC 35	29.563037	-11.357825	27.4	18	8.97	16	4	12	NIL	NIL	10	2.2	<0.002
R2 28 N	29.563212	-11.355635	40	25	10.1	NIL	NIL	>100	28.1	7.15	NIL	NIL	0.214
MC 8 HM	29.560759	-11.353165	NIL	NIL	10.2	NIL	NIL	0	NIL	NIL	NIL	NIL	0.309
R2 13N	29.939383	-11.196761	25	17	11.4	NIL	NIL	88	24.6	6.6	NIL	NIL	0.154
MC2 HM	29.824023	-11.454569	NIL	NIL	11.6	NIL	NIL	0	NIL	NIL	5	1.09	0.139
R2 15N	29.913756	-11.174697	26	17	11.6	NIL	NIL	>100	26.4	8.4	NIL	NIL	<0.002
MC 9HM	29.56106	-11.3491	NIL	NIL	18.2	NIL	NIL	0	NIL	NIL	NIL	NIL	0.164
MC 11HM	29.565979	-11.335845	NIL	NIL	19.3	NIL	NIL	0	NIL	NIL	NIL	NIL	0.151
MC 15HM	29.564466	-11.377271	25	16	37.2	NIL	NIL	0	23.9	6.73	NIL	NIL	0.018
MC 13HM	29.552831	-11.444636	NIL	NIL	67.2	NIL	NIL	0	NIL	NIL	NIL	NIL	0.228

Appendix IV: Field Geometry Points of Water Quality Parameters' sampling points used in Assessing the Relationship Between LULC and Water Quality.

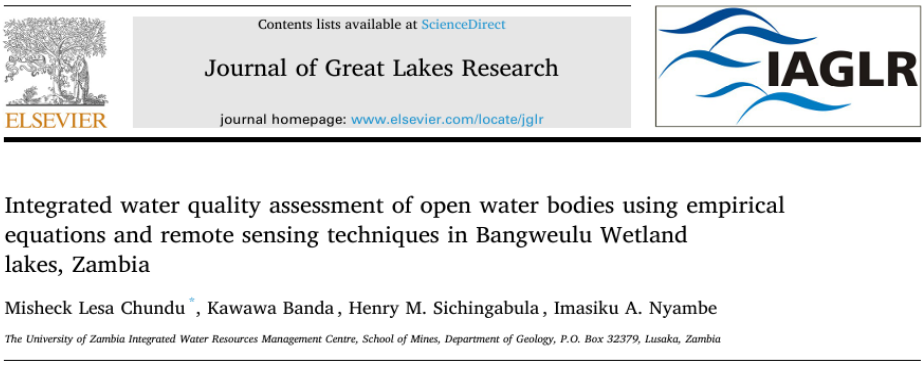
Sample Geometry Points		
Sample ID	**X**	**Y**
Chisangwa 1	29.98409	-13.2787
Chisangwa 2	29.98928	-13.2617
JK001	30.00064	-13.28713
JK002	29.98409	-13.2787
JK003	29.98928	-13.26174
JK005	30.29549	-12.55524
JK007	30.39089	-12.54286
JK008	29.74507	-11.4708
JK009	29.82541	-11.45503
JK011	28.738	-11.89011
JK012	28.74264	-11.76544
Kasongole River	29.52776	-11.2625
Mansa River	28.88894	-11.1934
MC 1 HM	29.751152	-11.46645
MC 11HM	29.565979	-11.335845
MC 12HM	29.570389	-11.325786
MC 15HM	29.564466	-11.377271
MC 8 HM	29.560759	-11.353165
MC 9HM	29.56106	-11.3491
MC14	29.564999	-11.455484
MC15	29.56447	-11.377288
MC16	29.554034	-11.332373
MC17	29.527936	-11.261939
MC18	29.541031	-11.231287
MC19	29.581376	-11.175625
MC20	29.619825	-11.158115
Mpanta	29.82499	-11.4548
Mukuku Bridge	29.8527	-12.1162
Mulembo Bridge	30.39086	-12.5429
Musamfwe	30.00054	-13.2896
Nkulumashiba Bridge	30.39393	-12.6207
R2-20	29.84573	-11.071022
R2-23	29.875994	-11.097429
R2-28	29.563214	-11.355632

Appendix V: Publications

1. Chundu, M. L., Banda, K., Lyoba, C., Tembo, G., Sichingabula, H. M., & Nyambe, I. A. (2024). Modelling land use / land cover changes using quad hybrid machine learning model in Bangweulu wetland and surrounding areas, Zambia. Environmental Challenges, 14(February), 100866. https://doi.org/10.1016/j.envc.2024.100866



2. Chundu, M. L., Banda, K., Sichingabula, H. M., & Nyambe, I. A. (2024). Integrated water quality assessment of open water bodies using empirical equations and remote sensing techniques in Bangweulu Wetland lakes, Zambia. Journal of Great Lakes Research, March, 102451. https://doi.org/10.1016/j.jglr.2024.102451



3. Chundu, M. L., Banda, K., Sichingabula, H. M., & Nyambe, I. A. (2025). Assessing land-use/land-cover influence on surface water quality using a weighted inverse distance function in Bangweulu sub-catchment area, Zambia. Physics and Chemistry of the Earth, 137(November 2024), 103813. https://doi.org/10.1016/j.pce.2024.103813